

Pred. No. is the number of results predicted by chance to have a

Burzio, L.

repeat_region
gene

TITLE Direct Submission
 JOURNAL Submitted (05-JUL-1993) Yoshiya Yoshida, Kanagawa Prefectural
 Public Health Laboratory; 52-2 Nakao-cho, Asahi-ku, Yokohama,
 Kanagawa 241, Japan (Tel:045-363-1030, Fax:045-363-1037)
 FEATURES Location/Qualifiers
 source 1..537
 /organism="Rickettsia japonica"
 /sub_species="YH"
 /db_xref="taxon:35790"
 /note="propagated in Vero cells"
 CDS 58..537
 /note="unnamed protein product"
 /codon_start=1
 /transl_table=11
 /protein_id="BAA03965.1"
 /db_xref="GI:1435056"
 /translation="MKLSIKMIIALATSMLCACNCPGGMNKQGTGYLLGAGAGALLIG
 SPGKGTQGLVGGVAGDGLGAVLGCGIGAGMDERLLAELT SQRLAETAPSGSNM"
 /rdpndng="GVYTPNFKYRNSTGQCYCREYTVVIGKKQKRAYGNACRPDQGMQVNV"
 BASE COUNT 176 a 92 c 129 g 140 t
 ORIGIN

| BASE COUNT | 176 a | 92 c | 129 g | 140 t | | Query Match | 9.6% | Score 46.6; | DB 1; | Length 537; |
|------------|-------|--|-------|-------|--|-----------------------|-----------------|-----------------|-----------|-------------|
| ORIGIN | | | | | | Best Local Similarity | 47.2%; | Pred. No. 0.62; | | |
| | | | | | | Matches 176; | Conservative 0; | Mismatches 194; | Indels 3; | Gaps 1; |
| QY | 100 | GGTCGGGTGTGGCGGCTGTTCGCCGCCAGCTGTTCCGTTAAAGGCTCTGGTCGTGTCGTCG 159 | | | | | | | | |
| DB | 163 | GGCGGTGCTGGTGGCGCAATTACTTGGTTCCTCAATTCCGTAAGGCACAGGACAGCTTGTG 222 | | | | | | | | |
| QY | 160 | ATGCCCATCGCGCTGCGGTTCTCGGCGGCTGATTTGGCTCTAAATCGGTCAGAGCATG 219 | | | | | | | | |
| DB | 223 | GGAGTAGGTGTAGTGCATTACTTGGAGCAGCTTCTTGGTGGACAAATCGGTCGAGGTATG 282 | | | | | | | | |
| QY | 220 | GACCAGCAGGATAAATCAAACTGAACAGCTCTGGGAAAAGTGAAGCGCGCCAGGTT 279 | | | | | | | | |
| DB | 283 | GATGACGAGGATGAAGAGCTTGCAGAGCTTACCTCACAGAGAGGTTTA---GAAACAGCT 339 | | | | | | | | |
| QY | 280 | ACTCGTTGGCGTAATCCGGACACCGGTAAACAGCTACTCTGTGGAAACCGGTTCCGACCTAC 339 | | | | | | | | |
| DB | 340 | CCTAGTGTAGTAACTGAGTAAATCCGGATACGCGCAATTACGGTTACGTACAC 399 | | | | | | | | |
| QY | 340 | CAGCGTTTACAACAACAGGAACGCGGTGAGCAGTACTGCGCGGAATTTACAGCAGAAACC 399 | | | | | | | | |
| DB | 400 | CCTAATAAACTTTAGAAAATGAGCACTGGTCAATATTGCGGTGAGTACACTCAACAGTT 459 | | | | | | | | |
| QY | 400 | ATGATCGCAGGTGAGAAACAGGAAATCTACGGCACCGGTCGCCCTCAGCCGGATGGCGCG 459 | | | | | | | | |
| DB | 460 | GTAATAGCGGGAACACACAAAAGCATACGGTATGCGTGGCCCAACCTGACGGACAA 519 | | | | | | | | |
| QY | 460 | TGGCAGGTGATTA 472 | | | | | | | | |
| DB | 520 | TGGCAAGTTGTGA 532 | | | | | | | | |

| | | | | |
|------------|---|------------------|---------|-----------------|
| RESULT 4 | 237523 bp | DNA | linear | BCT 11-NOV-1998 |
| RPXX04 | | | | |
| LOCUS | Rickettsia prowazekii strain Madrid E, | complete genome; | segment | |
| DEFINITION | 4/4 | | | |
| ACCESSION | AJ235273 | AJ235269 | | |
| VERSION | AJ235273.1 | GI:3861237 | | |
| KEYWORDS | complete genome. | | | |
| SOURCE | Rickettsia prowazekii. | | | |
| ORGANISM | Rickettsia prowazekii | | | |
| | Bacteria; Proteobacteria; alpha subdivision; Rickettsiales; | | | |
| | Rickettsiaceae; Rickettsiae; Rickettsia; typhus group. | | | |
| REFERENCE | 1 (bases 1 to 237523) | | | |
| AUTHORS | Andersson, S.G., Zomorodipour, A., Andersson, J.O., | | | |
| | Sicheritz-Ponten, T., Alsmark, U.C., Podowski, R.M., Naslund, A.K., | | | |
| | Eriksson, A.S., Winkler, H.H. and Kurland, C.G. | | | |
| TITLE | The genome sequence of Rickettsia prowazekii and the origin of mitochondria | | | |

JOURNAL Nature 396 (6707), 133-140 (1998)
 MEDLINE 99039499
 PUBMED 9823893
 REFERENCE 2 (bases 1 to 237523)
 AUTHORS Andersson, S.G.E.
 TITLE Direct Submission
 JOURNAL Submitted (11-NOV-1998) S.G.E. Andersson,
 Stiv. Andersson@molbio.uu.se, Dept. of Molecular Biology, University
 of Uppsala, Husargatan 3, Uppsala, S-751 24, SWEDEN
 Location/Qualifiers
 1. .237523
 /organism="Rickettsia prowazekii"
 /strain="Madrid E"
 /db_xref="taxon:782"
 164. .1123
 /gene="RP701"
 164. .1123
 /gene="RP701"
 /codon_start=1
 /transl_table=11
 /product="unknown"
 /protein_id="CAAL5137.1"
 /db_xref="GI:3861238"
 /translation="MAEPIITFLLFGNKRDIIDRINOKALNKLHIIPLOVNKESHIS
 DKERIFSEILKSYKECKPIEFNIOISNNVVKPIEPAQDLINAQNLNKTITFDKYN
 ALPQSELSKYWKQIIIEKVDHVFFNEADONLSIADGVPKDKTITIDIRLVTSFVN
 NLIYHETNHSUSSTIYKVDKLIKTAKNQDSKEIETWPIISVDEATNLITAKFGITS
 EDQIYGLKEIIEHLDPNNAENFKYVQISROFKKDLKDKRKHDSNFENTQKVI
 NDSNKDIQVEHITTSYVDQKEQKQPGELTRIFNFKNIVKQ"
 2074. .3348
 /gene="RP702"
 2074. .3348
 /gene="RP702"
 /codon_start=1
 /transl_table=11
 /product="unknown"
 /protein_id="CAAL5138.1"
 /db_xref="GI:3861239"
 /translation="MTTILVIVIIIMISLSALFAETATATASSPGKTHKIKIAGNKR
 AKTVLEVKREKVITLLIGNSLNITICTIATTFISLGNNGTIVASGVMAFIII
 VFAEVVPAIAVPEQTLTKMSSTIVIFLKLFRPINTALDYITKFCIFIRNLNPO
 ISGTEEVGEIHYHGGGVYKSERNMLGGILDINMTVSEIMTHRSNI IALNIDLPA
 ELIIKTLSSGAHTRIPLOMQRNDNIIGILNKLDKLALYENNDKKVDIINTLLTPK
 FIPDNALVVDOLHAFERNRNHFAVVDVEYGLIGITLEDVIEETVCGPITDEBRLNN
 ETIOESNTEFIKGTITRDINRELNDLNSBEDANTIAGLIHIIKARIIPNOGEVIKIP
 NFKIILIAKIANKIDSVKITVLPMEETISSE"
 3341. .5353
 /gene="RP703"
 3341. .5353
 /gene="RP703"
 /codon_start=1
 /transl_table=11
 /product="CYTOCHROME C-TYPE BIOGENESIS PROTEIN CCMF
 (ccmf)"
 /protein_id="CAAL5139.1"
 /db_xref="GI:3861240"
 /translation="MSKIGNFLLLTMCLSGILLILLIIPWLDYLIKKHRRNICLKFF
 PWICTSQWNTQATIFYASSLCSIFACTLVIATITDSFVONFLSHSTLKLPIYKI
 SGSSSHSGMMLWFCLLIQIVSCCYIFCLEDKMLKFSIIILSAIQLFSSFIYTSN
 PFMFSPVKGELGNLPMLODIALSIHPPLLYLGVSVAVPTIVCAIFLIVASRLD
 NLPFYFTGYKITQISDRITSRLLKIPAGCGILFTTIGISLGSWAWREYNWGLWF
 FDPVENISLFPWLSGIMLHSHIIVTTRNMQHWTIILSIITPLLVIFSTFLVBSGT
 NSIHAFASRERWYLLVILCIGIGLYLVIIRLNGSIEKTIYNTNWNTHSCNMT
 IKEKALIGNIFPFLSILLICILAIPTIYALFNDKPIIINEEFFIKFIIFVIPILC
 TIGFTNSSKKHILMLLSLIITYFILNKFVSILITIMTSIFLMLHNHYLLV
 KTNFRNSLKASSTSMILGHFVALIAFISPTNLSLQSEMDFIGEVGSKTFNAKVT
 LQNTKFAQKNYRQIAEFLWLEDHKNNTIILKPNRFYITIOKLSQSDTHSYLLYL
 YAVLSNIDGKI IIAKIVYKPMWSFIWIGVITLTAGFVIALIRKNSY"
 complement(7264. .12195)
 /gene="RP704"
 complement(7264. .12195)
 /gene="RP704"
 /codon_start=1
 /transl_table=11

/transl_table=11
 /product="CELL SURFACE ANTIGEN (sca5)"
 /protein_id="CAAL5140.1"
 /db_xref="GI:3861241"
 /translation="MAQPNFLPKKIIISALVLTASTATIVAGFSGVGAAMQAYNRRTTN
 AAATFDGIDGDAAGANIPVAPNSITANANNPLTFNPNHLSNLSFLDYANDLAVT
 INEDTTLGFIINIAQQAQAFNFNTVAAGKILNITGGITVQVQASNTINAOALTKVHGG
 AAILNANDLSGLSITFAAPSPVLEFLNIPPTQAPLPLGANSKIYVNGGNTLNTING
 FIQVSDNTFAGIKTINIDDCGLMFNSTPDAANTLNQVGGNTINFGIDTGKLVLY
 SKNGAATFNVITGLGKGIIEELNATAVAGKLIISOGGAANAVITGDTNAGRAAGFI
 VSDVONAAATISQVYAKNMVIOQANAGGVTFEHIYVDVGLGGTNTFTKADSKVIITE
 NSFGSTNFGNLDQIVVPDKTKILKGNFIDGVKNNGNTAGTIVTFNANGALVASTDPN
 IAVTNINAIEAGAGVVELSGIHLAEELRNGGSLFKIADCTVINGPNONALMNNNA
 LAAGSILQDLSAIITGIDGNGVNAALQHITLANDASKILADGANIIGANVGAIRH
 QANGSTIKLTNTONNIVYVFDLITDKTGIVDASSLNNTQTLTNGSIGTVVANTKT
 LAQNLGSSKILNAGDVAINELVIENNGSVOLAHNTLYLITKTIINAAQGGIIVAADP
 LNTNTLADGTLNLSAENPLSTIHFATKANADSLINVGKGNLYANNITINDANVS
 LHFSGGTSIVSGTGGQGHKLNNLILNDGTVFKLGDITFNGGTIEGKSIQISN
 NYTTHVESADNTGTLEFVNTDPTITLKNQAGYFGLKVIISGPGNIVFNEIGNV
 IVHGTAAINSISFENASLCTSLPSCPTPLDVTIKSTVGNCTVDENAPIVVSIDS
 MINNGIIGDKKIIILSLGSDNITVYNNATLYSGIRTKNNOCTVTLSCGMPNPGT
 IYGLGELNGSPKQVTFETDYNLGSIIANNVINYVTLTGGIAGTDFDAKITLG
 SVGNANRVFDSFSDPSMIVATQANKGTIVYLNALVSNISGLDTPVASTFTGN
 DSGAGLQGNIIYSONIDFGTYNLTILNSNVLGGGTTAINGEIDILLNNLIPANGTSTW
 GDNTSITTLNVSSNGITGVVIAEDAOVNATTTGTTIKIQDNANANFSGTQAVTLIQ
 GGARFNGTILGAPNEFAVTSNIFVKYELIRDSNODYVLTRTNDVNVVTVAGNSAIAN
 APGVSONTSRCLSTNTAAYNMMLLADPSDVAEFGAIATDTSAAVTVNLTQKT
 QDLLSNRLGTLRYLSNAETSIVAGSATGAVSSGDEAEVSGVWAKPFYNAQDKKGG
 TAGYKAKTGGVGVLDLNASNLMAIGAIGITDKHDDYKDKGDKTDINGLSFLYG
 SQQLKNFFAGNAIFTLNKVKRSQRYFESNGKMSKQIAAGNYDNMTFGNLIIFYG
 DYNAMPNVLTPMAGLSYLSKSNENYKETTGTVANKRINSKFSRDLIVCAKVGST
 VNIITDIVIPEIHSFVVKHNGKLSNQSMDLGTGTAPFISQPDRTAKTNSYNIGLSANI
 KSDAKMEYIGYDFNSASKYTAHOGTLKRVNF"
 complement(12953. .13216)
 /gene="RP705"
 complement(12953. .13216)
 /gene="RP705"
 /note="Pseudogene that lacks ATG or GTG initiation codon.
 Homology to quanosine-3,5-bis(diphosphate)
 3-pyrophosphohydrolase."
 complement(13563. .14507)
 /gene="RP706"
 complement(13563. .14507)
 /gene="RP706"
 /codon_start=1
 /transl_table=11
 /product="unknown"
 /protein_id="CAAL5141.1"
 /db_xref="GI:3861242"
 /translation="MKIRQAKPVIIGISPELTAERELFEHNPLIGLILFRNIRK
 NEKGQDKAALIKLIADIKVELGDNTIIADQEGMVKRLIAPFYDAPAAQTTELO
 TCYNSTIAKELREVGINLDFADVADLIHDGADKIIISDRSFGKEPEIVVPLISATA
 GLOEKYVTACIKHPIGHGRATVDSHIELPIDTSLKLTLEDTFKVKELAKYDIKILA
 MTAHVITLALDPDNPATLSQIVDYSKNSIFKGLIISDIETEMQALSQSGSMADITKGA
 DAGDVILECTGCFGNSEVIGSVASINDKFSDLFIS"
 complement(14552. .15025)
 /gene="RP707"
 complement(14552. .15025)
 /gene="RP707"
 /codon_start=1
 /transl_table=11
 /product="unknown"
 /protein_id="CAAL5142.1"
 /db_xref="GI:3861243"
 /translation="MINKTKVYVASEVIRIINTALHKLRYLETKIPGLSHYKNN
 RKYVTHDIDLLQKFLFVSRREYLEVGRHRSKDYLANFNSIHACSDTSRQNEFQG
 KNSCEILNLHVQQRQSSQSAVLSLMDMTVSSRIDILLTNFNLSTQIKILLIS"
 complement(15220. .15507)
 /gene="RP708"
 complement(15220. .15507)
 /gene="RP708"
 /codon_start=1
 /transl_table=11


```
RESULT 6
RSU76907
LOCUS
DEFINITION Rickettsia sp. 17kDa common-antigen gene, partial cds.
ACCESSION U76907
VERSION U76907.1 GI:2894794
SOURCE
ORGANISM Rickettsia sp.
REFERENCE
Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
Rickettsiaceae; Rickettsiae; Rickettsia.
1 (bases 1 to 448)
AUTHORS Davis, M.J., Ying, Z., Brunner, B.R., Pantofila, A. and Ferwerda, F.H.
TITLE Rickettsial relative associated with papaya bunchy top disease
JOURNAL Curr. Microbiol. 36 (2), 80-84 (1998)
MEDLINE 98087556
PUBMED 9425244
REFERENCE
2 (bases 1 to 448)
AUTHORS Ying, Z. and Davis, M.J.
TITLE Direct Submission
JOURNAL Submitted (01-NOV-1996) Tropical Research and Education Center,
University of Florida, 18905 SW 280 Street, Homestead, FL 33031,
USA
FEATURES
source
1. .448
/organism="Rickettsia sp."
/db_xref="taxon:789"
/note="Puerto Rico"
<1..>448
/note="putative"
/codon_start=3
/transl_table=11
/product="17 kDa common-antigen"
/protein_id="AAC02809.1"
/db_xref="GI:2894795"
/translation="ALAAASMLEACQPGPMKQSGTLLGGVLGVSGQGGTGRL
AavgAGALGAILNQIGAGMDQDEQKLAELTSORALEAPSGSSQWNPDPNGNYGT
VTPSKAYKNCTQCYREYTTQTVVVGKQKQKAYGTACRQPDGQWQV"
BASE COUNT 135 a 81 c 120 g 112 t
ORIGIN
Query Match 8.6%; Score 41.6; DB 1; Length 448;
Best Local Similarity 47.7%; Pred. No. 10;
Matches 184; Conservative 0; Mismatches 199; Indels 3; Gaps 2;
QY 87 TGGCGGCGCACCGCTGCGGTGTGGCGCGGTGTGGCGCGCAGCTGTTCGGTAAAGGCTC 146
Db 65 TTGTACGCTAATTGGCGGAACCTTTAGGAGGACTTGTGGTTTCAAAATTTGGTGTGTGATC 124
QY 147 TGTGCTGTGTCATGGCCATCGCGGTCTGGCGGTCTGGCGGTCTGATGGCTCTAAAT 206
Db 125 GGGACGACTTGTGCGGTAGGACGAGTGGTGTACTTGGTGCATCTTGGCAACCAAT 184
QY 207 CGGTACAGCATGACGACGAGGATCAAACTGAACCTGCTCTGCAAAAGTGAA 266
Db 185 TGTGTGAGGTATGATGAACAAATGAAACCTTGAGAGCTTACTTCTCAAGAGCTTT 244
QY 267 AGCGGCGCAGGTACTCTGTTGGCGTAAATCCGGACACCGGTACAGCTACTCTGTGAACC 326
Db 245 AGCTGCACCAACGGGTAGCAGTGTACAATGGCGTAACTCTGATATGTTATGATG 303
QY 327 GGTTCGACCTACCGGTTTACAACAAAGGACCGCGTACGACGTACTGTCGCGCAATT 386
Db 304 GTACTGTGACCAACGAAGAGCTTATAAAATAATACCG--GTCAATATTGTGCTGAATA 361
QY 387 TCAGCAGAAAGCCATGATCCAGCTCAGAACAGGAAATCTAGGCAACCGCTGCCCTCA 446
Db 362 CACCAAAACAGTTGTAGTAGTGGAAACAAACAAAGAGCTTATGGCACTGCCCTGCCGTCA 421
QY 447 GCCGATGGCCGCTGGCAGGTGATTA 472
Db 422 ACCTGATGGACAATGGCAACTTGTA 447
```

RESULT 7

```
AL646061
LOCUS
DEFINITION Ralstonia solanacearum GMI1000 chromosome, complete sequence;
segment 5/19.
ACCESSION AL646061
VERSION AL646061.1 GI:17427781
SOURCE
ORGANISM Ralstonia solanacearum.
Bacteria; Proteobacteria; beta subdivision; Ralstonia group;
Ralstonia.
1 (bases 1 to 198050)
AUTHORS Salanoubat, M., Genin, S., Artiguenave, F., Gouzy, J., Mangenot, S.,
Arlat, M., Billault, A., Brottier, P., Camus, J.C., Cattolico, L.,
Chandler, M., Choise, N., Claudel-Renard, C., Cunnean, S., Demange, N.,
Gaspin, C., Lavie, M., Moisan, A., Robert, C., Saurin, W., Schlex, T.,
Siguier, P., Thebault, P., Whalen, M., Wincker, P., Levy, M.,
Weissenbach, J. and Boucher, C.A.
TITLE Genome sequence of the plant pathogen Ralstonia solanacearum
JOURNAL Unpublished
REFERENCE
2 (bases 1 to 198050)
AUTHORS Boucher, C.A.
TITLE Direct Submission
JOURNAL Submitted (05-DEC-2001) Genoscope and CNRS UMR-8030, 2 rue Gaston
Cremieux, CP5706, 91057 Evry Cedex, France, Laboratoire de Biologie
Moléculaire des Interactions Plantes-Microorganismes INRA-CNRS,
BP27, 31326 Castanet-Tolosan Cedex, France, Fondation Jean
Dausset-CEPH, 27 rue Juliette Dodu, 75010 Paris, France, LMGM CNRS
118 Route de Narbonne, F 31062 Toulouse Cedex 4, Genoscope and INRA
URGV, 2 rue Gaston Cremieux, CP5706, 91057 Evry Cedex, France,
Laboratoire de Biometrie et Intelligence Artificielle INRA, BP27,
F31326 Castanet-Tolosan Cedex. Laboratoire de Genetique Cellulaire
INRA, BP27, F31326 Castanet-Tolosan Cedex
Christian Bouchet@toulouse.inra.fr
http://sequence.toulouse.inra.fr/R.solanacearum.html.
FEATURES
source
1. .198050
/organism="Ralstonia solanacearum"
/strain="GMI1000"
/db_xref="taxon:305"
151..642
/note="RS05084"
151..642
/note="RSC0770"
/EC_number="1.-.-.-"
/function="small molecule metabolism"
/note="Product confidence : putative
Gene name confidence : hypothetical
Predicted by Codon_usage
Predicted by Framed"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="PUTATIVE TRANSMEMBRANE DEHYDROGENASE (SMALL
SUBUNIT) OXIDOREDUCTASE PROTEIN"
/protein_id="CAD14472.1"
/db_xref="GI:17427782"
/translation="MHRPPSSDGDPAITRRQWLQSLAATAASLVGSVTLYAVAQ
ARPEPLDAFMGLSALDRAYGARLLAALQAQASAGFAAQLRPLAQSLAAGTLS
GLQQETALRIQAWTGVVDGTVVVTEQALMTYGVSDILVIRTVCNRPFGWAEPPVE
KOS"
642..2261
/note="RS05083"
642..2261
/note="RSC0771"
/EC_number="1.-.-.-"
/function="small molecule metabolism"
/note="Product confidence : putative
Gene name confidence : hypothetical
```


RESULT 8

AL646079

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

gene

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

gene

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

gene

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

gene

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

gene

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS 4400..5725
/gene="czcc"
/function="cell processes; protection responses"
/note="Product confidence : probable
Gene name confidence : probable
predicted by Codon_usage
predicted by Homology
predicted by Framed"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="PROBABLE COBAL-T-ZINC-CADMIUM OUTER MEMBRANE
RESISTANCE PROTEIN"
/protein_id="CAD17642.1"
/db_xref="GI:17430961"
/translation="MVTHTPCIRAAVALSCALLVRSGLAAAPAGEAASDTPAAPAQS
APASVGAHERMTLQQAIDRALATHPALAAAGYAVRASGAVDAQSLRNPAAFT
EDVRDRATMTAOLNPLELGGKRAARTRAELARDAGODAVALAELRADVRAFP
ELSTAQERLRLAQSADALAAQAEAAVRRVQAKVSPVEQTKAAVARAGAIELRDAQ
ADLLVAQRLSAFAGGAVAPPEADGSAETLPVPEAHRVLAATAAEPVALASTPRA
LRLALEIHRQALVNVSRKVPDVTLSVGAQRDSGANANNAVVGIAVPLFLDRNOG
SLLEAQLADKAEDYRALRLEQAALSDQVARLEAARQAVGALRQDVLPGAGRYDA
ARIGFEAGKFNFLVDLDAQRTLFQARAQYLAALSRAHQAAAGIDRILGR"
5752..6966
/gene="czcb"
/note="RSP0492; RS00370"
5752..6966
/gene="czcb"
/function="cell processes; protection responses"
/note="Product confidence : probable
Gene name confidence : probable
predicted by Codon_usage
predicted by Homology
predicted by Framed"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="PROBABLE COBAL-T-ZINC-CADMIUM RESISTANCE (CATION
EFFLUX SYSTEM PROTEIN) TRANSMEMBRANE"
/protein_id="CAD17643.1"
/db_xref="GI:17430962"
/translation="MTLPQORRAIALIVLAGALAAALWGMGARAQAAHAEAEOP
AAHAEHAQAPKAEAAEPHETIAMTAEAIANAALGVETAGALQRTVEVLPGEIRLN
DDRTAHVVRVAGVAEQVAVELQPVTLGQALLAVLSSPAGSDQSRSELOAQRALAQ
EAHAREKRLWFOGIAAQQOQARTLAQEAIRADNARQKLAIIAGAPAGPALNRFEL
RAPEDGVVVAHRLSQGEAAVQAEAAFTIADRTYMADFAVTAKDLAEVTHATATVRA
TATGTAVQGVKYVYGALLGEQTSAPARVTLDNPKRAWRPMGFVSVSVEGRVPPVA
VRADAVOTVDEKPSVEFVPPVAGFEVRPVKTRGSDGTWTLELDLPLAGTRYAANAAYVL
KAEAGKSADHGH"
6977..10171
/gene="czca"
/note="RSP0493; RS00371"
6977..10171
/gene="czca"
/function="cell processes; transport of small molecules"
/note="Product confidence : probable
Gene name confidence : probable
predicted by Codon_usage
Query Match 8.4%; Score 40.8; DB 1; Length 213050;
Best Local Similarity 58.1%; Pred. No. 17;
Matches 72; Conservative 0; Mismatches 52; Indels 0; Gaps 0;
Qy 85 GTTGGCGGGCCACCGGTGGCGGTGTGGCGGCTGTGGCGGACAGCTGTTCGGTAAAGGC 144
Db 20248 GTGGCGCGGTCTGCGCGGCTGTGGCGGCTGTGGCGGACACAGATCGGTGGCGGG 20307
Qy 145 TCTGTCTGTCTCATGCCATCGCGGCTGGCGGTCTGGCGGCTCTGATGGCTCTAAA 204
Db 20308 CCGCGCAATACGGTGGCCACCATCGCGGCGGTGTGGCGGCGGCTGTGGCGGCAACACC 20367
Qy 205 ATCG 208
|||||

Db 20368 ATCG 20371
RESULT 9
RIRANT17KA
LOCUS RIRANT17KA 539 bp DNA linear BCT 26-APR-1993
DEFINITION R.conorii 17K genus-common antigen gene, complete cds.
ACCESSION M28480
VERSION M28480.1 GI:152457
KEYWORDS antigen.
SOURCE R.conorii DNA.
ORGANISM Rickettsia conorii
Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
Rickettsiaceae; Rickettsia; spotted fever group.
REFERENCE 1 (bases 1 to 539)
AUTHORS Anderson,B.E. and Tzianabos,T.
TITLE Comparative sequence analysis of a genus-common rickettsial antigen
gene
J. Bacteriol. 171 (9), 5199-5201 (1989)
MEDLINE 89359171
PubMed 2768201
COMMENT Draft entry and printed copy of sequence for [1] kindly submitted
by B.Anderson, 27-OCT-1989.
FEATURES
Location/Qualifiers
1..539
/organism="Rickettsia conorii"
/db_xref="taxon:781"
1..6
-35_signal
21..27
-10_signal
38..>539
/product="17 kD mRNA"
60..539
/codon_start=1
/note="17 kD antigen precursor"
/transl_table=11
/protein_id="AAA26376.1"
/db_xref="GI:152458"
/translation="MKLLSKIMIIATLSMLOACNPGPMNKQGTGTLGGAGGALLG
SQFGKRGQLVGVGALLGAVLGGQIGAGDEQDRRLAELTSRALETAPSGSNVW
RNPDNGNYGVTPNKTVRNSTGQCREYTOTVTVGGKQKAYGNACRPDQGWVN"
60..119
sig_peptide
/note="17 kD antigen signal peptide"
120..536
mat_peptide
/product="17 kD antigen"
BASE COUNT 179 a 93 c 125 g 142 t
ORIGIN
Query Match 8.3%; Score 40.2; DB 1; Length 539;
Best Local Similarity 46.1%; Pred. No. 23;
Matches 172; Conservative 0; Mismatches 198; Indels 3; Gaps 1;
Qy 100 GGTGGCGGTGTGGCGGCTGTGGCGGCTGTGGTAAAGGCTCTGGTCTGTGTGCG 159
Db 165 GCGCGTGTGGCGGCGCATTTACTTGTCTCAATTCGTTAAGGCGCAAGGACAGCTTGT 224
Qy 160 ATGGCCATCGGCGGTGGCGGTCTGGCGGTCTGATTTGCTTAAATCGGTGAGAGATG 219
Db 225 GGAGTAGGTGTAGTGTGATTTACTTGGAGCAGTTCTTGTGGACAAATCGGTGAGTATG 284
Qy 220 GACGAGCAGGATAAATCAAACTGAACAGTCTCTGGAAGAGTGAAGCCGCCAGGTT 279
Db 285 GATGAACAGGATAGA---AGACTTCAGAGAGCTTACTCAGAGAGAGCTTTAGAAACAGCT 341
Qy 280 ACTCTTTGGCGTAAATCCGACACCGCTAATCAGCTACTCTGTGGAACCGGTTCCGACCTAC 339
Db 342 CCTAGTGTAGTAACTAGTAGTGGGTAAATCCGGATAACGGCAATTACGTTACCTAACA 401
Qy 340 CAGCGTTTACAACAACAGAGAGCGGCTGAGCAGTACTCGCGGCAATTCAGCAGAAAGCC 399
Db 402 CCTATAAACTTATAGAAATAGCAGTGTGCTCAATATTGCGGTGAGTACACTCAACAGTT 461
Qy 400 ATGATCGCAGGTTCAGAAACAGAGAAATCTACGCGACCGGCTGCCCTCAGCGGATGCCGC 459
|||||

Db 462 GTAATAGCGGAAACAAACAAAGGATACGGTAATGCATGCCGCCAAACCTGACGGACAA 521

QY 460 TGCAGAGTGATTA 472

Db 522 TGGCAAGTTGTGA 534

RESULT 10

RIRANTI17KD RIRANTI17KD 539 bp DNA linear BCT 26-APR-1993

LOCUS R.rickettsia 17K genus-common antigen gene, complete cds.

DEFINITION M28479

ACCESSION M28479.1 GI:152463

VERSION M28479.1

KEYWORDS antigen.

SOURCE R.rickettsia DNA.

ORGANISM R.rickettsia rickettsii

REFERENCE 1 (bases 1 to 539)

AUTHORS Anderson,B.E. and Tzianabos,T.

TITLE Comparative sequence analysis of a genus-common rickettsial antigen gene

JOURNAL J. Bacteriol. 171 (9), 5199-5201 (1989)

MEDLINE 89359171

PUBMED 2768201

COMMENT Draft entry and printed copy of sequence for [1] kindly submitted by B.Anderson, 27-OCT-1989.

FEATURES

source

1. .539

/organism="Rickettsia rickettsii"

/db_xref="taxon:783"

1. .6

-35_signal

-10_signal

mRNA

37. .>539

/product="17 kD mRNA"

60. .539

/note="17 kD antigen precursor"

/codon_start=1

/transl_table=11

/protein_id="AAA26379.1"

/db_xref="GI:152464"

/translation="MKLLSKIMIIALATSMLOACNGPGGNKOGTCTLLGGAGALLG

SQPKGKGLVGVGALLGAVLGGIGAGMDQRRLAELTSQRALETAPSSNVW

RPNPNNGYGVTPNKTIRNSTGQYCREYQTQVVGKQKAYGNACRQPDGQWVNV

60. .119

/note="17 kD antigen signal peptide"

120. .536

/product="17 kD antigen"

mat_peptide

179 a 92 c 125 g 143 t

BASE COUNT

ORIGIN

Query Match 8.3%; Score 40.2; DB 1; Length 539;

Best Local Similarity 46.1%; Pred. No. 23;

Matches 172; Conservative 0; Mismatches 198; Indels 3; Gaps 1;

QY 100 GGTGCGGTGTGGCGGTGTGGCGGCAGCTGTTCGGTAAGGCTGTGCTGTGTGTGTCG 159

Db 165 GCGGTGCTGGCGGCAGATTACTTGTTCATTCGGTAAGGCGCAAGGACAGCTTGT 224

QY 160 ATGCCCATCGCGGTGTGCGCGGTCTGTGCTCTAAATCGTCAAGCGTCAGAGCATG 219

Db 225 GGAGTAGGTGAGTGCAATTACTTGGAGCAGCTTCTTGGTGACAAATCGGTGCGAGGTATG 284

QY 220 GACCAGCAGGATAAATCAACACGAGTCTCTGGAAGAGTGAAAGCGCGCGAGGTT 279

Db 285 GATCAACAGGATAGA---AGACTTCAGAGCTTACCTCACAGAGAGCTTTAGAACAGCT 341

QY 280 ACTCGTTGGGTATTCGCGACACCGGTAAACAGTACTCTGTGGAACCGGTTTCGACCTAC 339

Db 342 CCTAGTGTAGTAACTGATGATGCGTAAATCCGGATAACGCAATACCGGTTACGTAACA 401

QY 340 CAGCGTTACAAACACAGGACGCGCTCAGCAGTACTGCCCGCAATTCACGACAAGCC 399

Db 402 CCTAATAAACTTATAGAAATAGCACCTGGTCAATATTGCCGTGAGTACACTCAAAACAGTT 461

QY 400 ATGATCGCAGGTGAGAAATCTAGCGACCGTCCCTCAGCGGATGGCCGC 459

Db 462 GTAATAGCGGAAACAAACAAAGGATACGGTAATGCATGCCGCCAAACCTGACGGACAA 521

QY 460 TGCAGAGTGATTA 472

Db 522 TGGCAAGTTGTGA 534

RESULT 11

AE008675

LOCUS Rickettsia conorii Malish 7, section 107 of 114 of the complete genome.

DEFINITION AE008675 AE006914

ACCESSION AE008675.1 GI:15620413

VERSION AE008675.1

KEYWORDS Rickettsia conorii.

SOURCE Rickettsia conorii.

ORGANISM Rickettsia conorii

REFERENCE 1 (sites)

AUTHORS Ogata,H., Audic,S., Barbe,V., Artiguenave,F., Fournier,P.E., Raoult,D. and Claverie,J.M.

TITLE Selfish DNA in protein-coding genes of Rickettsia

JOURNAL Science 290 (5490), 347-350 (2000)

MEDLINE 20485642

PUBMED 11030655

REFERENCE 2 (sites)

AUTHORS Ogata,H., Audic,S. and Claverie,J.-M.

TITLE Selfish DNA and the origin of genes

JOURNAL Science 291 (5502), 252-253 (2001)

REFERENCE 3 (bases 1 to 10127)

AUTHORS Ogata,H., Audic,S., Renesto-Audiffren,P., Fournier,P.-E., Barbe,V., Samson,D., Roux,V., Cossart,P., Weissenbach,J., Claverie,J.-M. and Raoult,D.

TITLE Mechanisms of evolution in Rickettsia conorii and R. prowazekii

JOURNAL Science 293 (5537), 2093-2098 (2001)

MEDLINE 21442074

PUBMED 11557893

REFERENCE 4 (bases 1 to 10127)

AUTHORS Ogata,H., Audic,S., Renesto-Audiffren,P., Fournier,P.-E., Barbe,V., Samson,D., Roux,V., Cossart,P., Weissenbach,J., Claverie,J.-M. and Raoult,D.

TITLE Direct Submission

JOURNAL Submitted (26-APR-2001) Unite des Rickettsies, Faculte de Medecine, CNRS UMR 6020, Universite de la Mediterranee, 27 boulevard Jean Moulin, Marseille Cedex 05 13385, France

COMMENT A public version of R. conorii genome database is accessible at <http://igs-server.cnrs-mrs.fr/>. The database intends to provide updated data. Annotation of the genome is an ongoing task whose goal is to make the genome sequence more useful. Comments to the authors are appreciated.

FEATURES

source

1. .10127

/organism="Rickettsia conorii"

/strain="Malish 7"

/db_xref="taxon:781"

111. .890

/gene="map"

/note="RC1276"

111. .890

/gene="map"

/codon_start=1

/transl_table=11

/product="methionine aminopeptidase [EC:3.4.11.18]"

/protein_id="AAL03814.1"

/db_xref="GI:15620414"

/translation="MTIKHTGKDFIKMRAAGKLAATLDFITDHYKPNVNTSLNCL

CHNFTSHNAVAPLNYKGFPSKICISINIVCHGIPNDRPLKNGDINVIDVILDG

WYGDTSRMVTVGDVAIKPKRLIQVTIDAMMKGIEVVRPGAKLGDIGHAIQSYAEKHNY

```
SVRDYTGCHGIGRVFHDPEPSILNYGRSGTGLTLEEGMFFTVPEPMINAGNYDTILSKLD
GWTYTRDKSLSAQFHTIGVTGDKGFEIFTLSPKKLDYPPY"
/notes="REP03, repeated element"
/translation="1064..1083"
/notes="REP02, RS3-like repeat"
1141..1452
/genes="RC1277"
/genes="RC1277"
/genes="RC1277"
/codon_start=1
/transl_table=11
/product="similarity to DNA repair protein (RadC)"
/protein_id="AAL03815.1"
/db_xref="GI:15620415"
/translation="MLENNLEFNSDEIDTPHYGHRKRVKRFVAAGAEYFSDYEL
LEVWLSAIPRKDIKPLAKKLLDHEIDTLNLDKERLLSKGTNENLYINFALRELI
N"
1457..1825
/genes="RC1278"
1457..1825
/genes="RC1278"
/codon_start=1
/transl_table=11
/product="similarity to DNA repair protein (RadC)"
/protein_id="AAL03816.1"
/db_xref="GI:15620416"
/translation="MKQIKINKNINASWSSWIDYLVKVMGNMRLEQFRILFLNKNIL
TAEVLSTGTTDOAAVYPREIKRALFNEASSLIIVHHHPSGPEPSKANIHTMNTKIY
EKQPINIIVNLICSYEK"
1812..2795
/genes="ray2"
/notes="RC1279"
1812..2795
/genes="ray2"
/codon_start=1
/transl_table=11
/product="undecaprenyl-phosphate
alpha-N-acetylglucosaminyltransferase [EC:2.4.1.1]"
/protein_id="AAL03817.1"
/db_xref="GI:15620417"
/translation="MNINSYSLTFISFTSTAILTYILTKVLTIGLVDPISNRSHD
KITRGGGLAIVFVVMIALSGEYITSNNFVNSIKLPLLLAIAISFEDDLKAVPIL
IRLVHLICAFAPFISOLNSHLIYISLVIAUSGFNIYFNMDFGIDGMSCEVSIH
LSSTMILCFQFTINNPYFTTSVNVIIIGSCGFLFNHHPAKIFLDGVSISLGF
LTGLCLLALTSTNLFACASLYIYTDVLTILIRLANKETIQPHLKHFFOKAV
QKGGSHKQWSTIAVCNIFLMIISVTSLSHFPVISIMLALAVITLTQSLK"
2792..3115
/genes="RC1280"
2792..3115
/genes="RC1280"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAL03818.1"
/db_xref="GI:15620418"
/translation="MKLIKEDDKIVRHIFAKQHPLLPEIMINWKNKIVGFNFSTKALP
LKITYTYKKQINTLFIAEDNATAALPYQDIILERIKIYLGFAAHQMNVTYFK
GKSL"
complement(3198..3227)
/notes="REP02, RS3-like repeat"
3313..3353
/notes="REP01, repeated element"
complement(3509..4258)
/genes="RC1281"
complement(3509..4258)
/genes="RC1281"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAL03819.1"
/db_xref="GI:15620419"
/translation="MKKLLIIAAASTALLTSLGSFADCDMNSSVDSSTNSMSSSVEN
```

```
OWYLKLNAGGVIEKTKPKGADEKLNNIKSNKISNTGFTGEIGAGYVIMDLNRLDILT
GTVASSHLKSKTYPPDGNSESVKKNKPIIVSVLLNGYVDFDLSEFKYFAGAGVGAAY
KEKIHSDIKGGVDTFNGTKNKNTFAYOLSLGTSFEVAGGVKAELVYSWRDYGKTK
NTTKINGDKVFKGGTHYKGNLMAGLRFDM"
4438..4463
/notes="REP03, repeated element"
complement(4505..5188)
/genes="RC1282"
complement(4505..5188)
/genes="RC1282"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAL03820.1"
/db_xref="GI:15620420"
/translation="MLTMKLLIIAATSAITMLSSVSFAEGMDHEWYLRIDTGAAMFN
EEKQATGVKLSKNTVPVALGIGYIYSENFADLTGLTIGGKLSGATNAPFTR
TNISASHKPTITRLLINGYVDTLTFNDFMDVPAGAGVGSALVKEKITYNGITGLSSNTK
NRTNWSKTLIGTSAQIADGVKVELAYSWIDDGRTKSNVIYQGTSVPTGGMHYQSHN
LTAGIRFDI"
complement(5338..5508)
/notes="REP06, A and T rich repeated element"
complement(5478..5828)
/genes="fdxa"
/notes="RC1283"
complement(5478..5828)
/genes="fdxa"
/codon_start=1
/transl_table=11
/product="ferredoxin"
/protein_id="AAL03821.1"
/db_xref="GI:15620421"
/translation="MTYVTVDECVKCYTDVCEVPCDFYEGEGEREDDFMLVINPDEC
IDCGVCPDCPIGAIKPESGLIEWERAKDFIENKGWKNITKKKTALPDADKFKDEK
DKFNKYIGSVKIAN"
complement(5833..6525)
/genes="ccmc"
/notes="RC1284"
complement(5833..6525)
/genes="ccmc"
/codon_start=1
/transl_table=11
/product="heme exporter protein C"
/protein_id="AAL03822.1"
/db_xref="GI:15620422"
/translation="MLKLLTPYFNKLSKVIPILAISTCVMIIIGLYLALIVSPDY
QQGFVRIMYVHVFPASNMALGVYFPMAYCSFYSYLWKTTSYLLAVASSYVGAFTLI
SLVTSGLMGKPIWGTWVWDARLTSMLILFLIYLSYIIIVNSADNRKQNPSSIAL
IGLINPIVKSFSVNIWYSLHQPASVLRGSPITHSSMLKPLIIMFISILYFLVLL
RTAILDKIKNR"
complement(6540..7046)
/genes="RC1285"
complement(6540..7046)
/genes="RC1285"
/codon_start=1
/transl_table=11
/product="cytosine deaminase [EC:3.5.4.1]"
/protein_id="AAL03823.1"
/db_xref="GI:15620423"
/translation="MNILLDEANFNFFMEQALQKQAFDKNFVPGVAVVDRLHQ
KIIASTHNTEKNALYHAETIAINEACNLISKNLNDYIYVTLPCAMCAAAIAH
SRKLRFYGASDSKHGVVESNLRYFNSSACFHRPEIYSGLIAEDSGLMKKEFKRIRT
VISHRM"
complement(7027..7947)
/genes="p34"
/notes="RC1286"
complement(7027..7947)
```

```
Query Match 8.3%; Score 40.2; DB 1; Length 10127;
Best Local Similarity 46.1%; Pred. No. 23;
Matches 172; Conservative 0; Mismatches 198; Indels 3; Gaps 1;
QY 100 GGTGCGGTTGTGGCGGTGTGTCGCCGCCAGCTGTTCCGTAAGGCCTGTCGTGCTGTCG 159
```


| | | | |
|------------|------------|---|----------------------------|
| Db | 8239 | GGCGTCTGGCGCGGCGATTACTTGGTTCTCAATTCGGTAGGCGCAAGACAGCTTGT | 8298 |
| QY | 160 | ATGCCCATCGCGCGTCTGGCGGCTGATTCGGCTTAAATTCGGTCAGAGCATG | 219 |
| Db | 8299 | GGAGTAGGTGAGTGCATTACTTGGAGCAGTCTTGGTGACAAATCGTGCGAGGTATG | 8358 |
| QY | 220 | GACCAGCAGATTAATCAACTGAACCATGACCTCTCTGGAANAAGTGAAGCGCGCAGGTT | 279 |
| Db | 8359 | GATGAACAGGATAGA---AGACTTGCAGAGCTTACCTTCACAGAGAGCTTTAGAAACAGCT | 8415 |
| QY | 280 | ACTCGTTGGCGTATCCGGGACACCGGTAAACAGCTACTCTGTGGAACCGGTTCCGACCTAC | 339 |
| Db | 8416 | CCTAGTGGTAGTAGTAGAATGCGGTAACTCCGATACCGGCAATTACGGTTAGGTAACA | 8475 |
| QY | 340 | CAGCGTTTACAAACAGGACCGCGTCAGCAGTACTGCCGCGAATTTACGACGAAAGCC | 399 |
| Db | 8476 | CCTAATAAACTTATAGAAATAGCAGTGTCAATATTGGCGTGAGTACACTCAACAGATT | 8535 |
| QY | 400 | ATGATCCAGGTGAGAAACAGGAATCTAGCGACCGCGTGCCTCAGCGGATGGCCGC | 459 |
| Db | 8536 | GTAATAGCGGGAACAAACAAAAGCATACGGTAATGATGATGCCCAACCTGACGGACAA | 8595 |
| QY | 460 | TGGCAGGTGATTA 472 | |
| Db | 8596 | TGGCAAGTTGTGA 8608 | |
| RESULT 12 | | | |
| AE004537 | | | |
| LOCUS | | | |
| DEFINITION | AE004537 | 13946 bp | DNA linear BCT 30-AUG-2000 |
| ACCESSION | AE004537 | | |
| VERSION | AE004537.1 | | |
| KEYWORDS | GI:9946960 | | |
| SOURCE | | | |
| ORGANISM | | | |
| REFERENCE | | | |
| AUTHORS | | | |
| TITLE | | | |
| JOURNAL | | | |
| MEDLINE | | | |
| PUBMED | | | |
| REFERENCE | | | |
| AUTHORS | | | |
| TITLE | | | |
| JOURNAL | | | |
| FEATURES | | | |
| source | | | |
| gene | | | |
| CDS | | | |

| | |
|--|--|
| /transl_table=11 | |
| /product="probable esterase" | |
| /protein_id="AAG04436.1" | |
| /db_xref="GI:9946961" | |
| /translation="MPRLSHRNLPMTPOGHCDRFPALAEAFARLFDNPOERGAALC | |
| LEVGSTVVDLGGVADKDEOPWODTILNLPSCKTFTTAVAVLQLVAGKLELDAP | |
| VAHYWEFAAGKAAISVQLLCHRAGLPALEQMPPEALIDMQMTTALAEPEPWT | |
| PEAHGYAPITYGWLIGEVIRRVDRGPEAIVARTAAPLGLDFHVLDDSDQFGRVAH | |
| MARSGSLGDAARMLKTMSEPLATTTAFNTPPSILTSTNKKPWRMRQOPAAANGH | |
| GNARLAGFYSGLRLLELGLQLAHEHVAVEDRTLLTRTFGLGTMLDQPGVAN | |
| ATVGLPKAFGHPGAGSGTGFADPERDLAFSEFVSNALGPVVLMDPRAQQLARVAGECL | |
| " | |
| 1453..2334 | |
| /gene="PA1048" | |
| 1453..2334 | |
| /gene="PA1048" | |
| /codon_start=1 | |
| /transl_table=11 | |
| /product="probable outer membrane protein" | |
| /protein_id="AAG04437.1" | |
| /db_xref="GI:9946962" | |
| /translation="MKLLKSTLLLCMTASGCSYFQTDASSQAPAKAPQASSHSW | |
| PFKDDQTVAKOATEOKAEAEHVSSHWMPFGKDEGGRDVAAKAEOKAEATAAKK | |
| EYAAAEKPAANPMWPFSGASSKASGKAPALVQOKVTKEWLDAHEKALREATAGSEF | |
| TIERNALIVITAPADYSNPKRHTMLPITLNPGLKVAQMAQDFQSLGLVGHGTHDS | |
| TGSKAVNDKLSFERAQSVAAIFRLSLGLKQDLKGVGSDMPRADNAAAGRAQNRV | |
| EVLYTQASLLSLAQN" | |
| 2433..3080 | |
| /gene="pdxH" | |
| /note="PA1049" | |
| 2433..3080 | |
| /gene="pdxH" | |
| /codon_start=1 | |
| /transl_table=11 | |
| /product="pyridoxine 5'-phosphate oxidase" | |
| /protein_id="AAG04438.1" | |
| /db_xref="GI:9946963" | |
| /translation="MTQSLADMREYTRDGLSEANAPSPFSLFRQWDDAVKTERLP | |
| VEFNAMTLATVDADGPHCRILLLKLDGERGFTFTTYESAKGRQLAANRAAMTFFW | |
| PALEROVRIEGSVKVTPEESDAYQVRPLGSLRGAWASQSRVIAADRLERLLADT | |
| ERRFADQPPSCPEHWGGYRLLPQRIEFWQGRPSRLHDLRDYRRQDGGWLRRLAP" | |
| 3088..4200 | |
| /gene="PA1050" | |
| 3088..4200 | |
| /gene="PA1050" | |
| /codon_start=1 | |
| /transl_table=11 | |
| /product="conserved hypothetical protein" | |
| /protein_id="AAG04439.1" | |
| /db_xref="GI:9946964" | |
| /translation="MFLDSTLAQHIIVDRAMAILPHNINVMQAQMIIGSGDPSRLHT | |
| RHEGAQLVLANRRVVIDEQAACLRGVRPNLPLHAERLVGVIGITGAPVVRPY | |
| AEIVRAAEMLMEOROMORORHEARLRLDLPCHGLGALAADAHEHLGLAMHW | |
| PQAVLELPENADPLPCQARLLAALGGRAEHWSALGERLVFCRPVPAHDTLEHL | |
| ERADARGWGLCARCGADPAHDLAELRASAARADLDYARSLRPGRRVYVLDGHRLAS | |
| LIFSQRDSQVQVQLAPLQRLDHPDQSLAATLESWCARDHGAQAQALGTHRNSL | |
| RYRLERIASISGLDRLARLHROQLLGLQLLVPRRG" | |
| 4359..5717 | |
| /gene="PA1051" | |
| 4359..5717 | |
| /gene="PA1051" | |
| /codon_start=1 | |
| /transl_table=11 | |
| /product="probable transporter" | |
| /protein_id="AAG04440.1" | |
| /db_xref="GI:9946965" | |
| /translation="MTVLILILIALIAFTVLSTRKLHPFLALLAALIAAGIAQVPT | |
| AEIVKTGTGFGSILGIVIGIVILGIIIGVILERSGAATMAESVILIGERFPTLTM | |
| SIIGYLVISIPFCDSGFVILNSLKNALAKMISTAMSVALTGLYATHTFVPPTPG | |
| PIAAGNLGLDASLGLVIVVGLVAVFTAMAGMWANRFVKDIPLVDDGVVQVTEED | |
| FSELRYKGLPSATCAFIEVPILLICLGSVAVTFPSKPLGEGVLFACLNFIQGPVI | |
| ALLVGLALACTLLKGSKREEFHVEVIGSQSAAPILLITGAGGAFGAMIKITPLGDI | |
| LGLTSLALGIGFMFVMAALKTAQGSTTVALVTTTVALVAPLLPQLGLDSEMGRLT | |

| | | | | | |
|--------------|---|--|--------------|--|--|
| misc_feature | 1. .15769 | | | | |
| | /note="basepairs 1-15769 are not integral part of BAC clone F7K2" | | | | |
| gene | 3. .821 | | gene | | /gene="AT4g22410" |
| gene | /gene="AT4g22380" | | CDS | | complement(join(6112. .6297,6675. .6812,6908. .7075, 7168. .7353,7429. .7513,7775. .8034)) |
| | complement(join(3. .98,594. .821)) | | | | /gene="AT4g22410" |
| | /gene="AT4g22380" | | | | complement(join(6112. .6297,6675. .6812,6908. .7075, 7168. .7353,7429. .7513,7775. .8034)) |
| CDS | complement(join(<3. .98,594. .821)) | | | | /gene="AT4g22410" |
| | /gene="AT4g22380" | | | | /note="similarity to snRNP assembly factor like protein, Homo sapiens, PID:g5730025" |
| | /note="strong similarity to NHP2/RS6 FAMILY PROTEIN, Homo sapiens, PID:g4826860 | | | | /codon_start=1 |
| | Contains Ribosomal protein L7Ae signature AA73-90 | | | | /product="putative protein" |
| | contains EST gb:F13981" | | | | /protein_id="CAB52815.1" |
| | /codon_start=1 | | | | /db_xref="GI:5738372" |
| | /product="Ribosomal protein L7Ae-like (fragment)" | | | | /translation="MVEQVLDHFERFCVSLSNLNVYACLVCGKYFGQRSGKSHAY |
| | /protein_id="CAB52812.1" | | | | THSLEAGHHVYINLLTEKYVCLDPSEINDPSLDDIRHVLNPRFSRAQVNELDKRWQ |
| | /db_xref="GI:5738369" | | | | SRAALDGDYLPQGVNGLNNIQKTEFVAVNTIQSLMRVTPLRNFHHPENYORCKSPLVHC |
| | /translation="MTGEVYNPKAYPLADSQLSITIMDLVOOATNYKOLKKGANEATK | | | | FGELTKIKHARNFQGVSPHEFLQAVMKRKRFRIGQOSDPVEFMSWLLNTLHMDL |
| | TLMRGISEFWMAADAEPLILLHLPLAEDKNVPYVFVPSKQALGRACGVTRPVIAC | | | | RKSDASSIIHKCFQGELEVVKEFOGNEKEISRMSEFLMGLDLPPLPPPKDVMENI |
| | SVTSNE" | | | | IPQVALFDLLKKFDGTEVTPRKLARMYRIVKSPRYLHFMHVRFKNNFFKRNKPI |
| | | | | | TIGESV" |
| exon | complement(3. .98) | | exon | | complement(6112. .6297) |
| | /gene="AT4g22380" | | | | /gene="AT4g22410" |
| | /number=1 | | | | /number=1 |
| Intron | complement(99. .593) | | Intron | | complement(6298. .6674) |
| | /gene="AT4g22380" | | | | /gene="AT4g22410" |
| | /number=1 | | | | /number=1 |
| exon | complement(594. .821) | | exon | | complement(6675. .6812) |
| | /gene="AT4g22380" | | | | /gene="AT4g22410" |
| | /number=2 | | | | /number=2 |
| gene | 1789. .2973 | | Intron | | complement(6813. .6907) |
| | /gene="AT4g22390" | | | | /gene="AT4g22410" |
| gene | complement(1789. .2973) | | | | /number=2 |
| | /gene="AT4g22390" | | | | complement(6908. .7075) |
| CDS | complement(1789. .2973) | | exon | | /gene="AT4g22410" |
| | /gene="AT4g22390" | | | | /number=3 |
| | /note="similarity to various predicted proteins, Arabidopsis thaliana" | | Intron | | complement(7076. .7167) |
| | /codon_start=1 | | | | /gene="AT4g22410" |
| | /product="putative protein" | | exon | | complement(7168. .7353) |
| | /protein_id="CAB52813.1" | | | | /gene="AT4g22410" |
| | /db_xref="GI:5738370" | | Intron | | complement(7354. .7428) |
| | /translation="MAECTPTDLINEMFLRATTLVKCRVLSKPCFSLIDSPEFVSSH | | | | /gene="AT4g22410" |
| | LRRRLTGHMLLGRRLRTVELDSPENVSDIPHLQAGGTFEFGSFGVIGLC | | exon | | complement(7429. .7513) |
| | NSPVDLAIFNPSTRKIRLPIEPDIPERDIREYVYGLGDSVGDDKFKVRIVOCK | | | | /number=5 |
| | LKEGRKKFPCPEVEKVPISLKMSKRWKRMFEPQILWISYYHLLPRGYGVVNNHL | | Intron | | complement(7514. .7774) |
| | HWLPRQGVIAFNAIKYDLASDDIGVLSFPQELYIEDNMDIGVLDGCVLCMYDEY | | | | /gene="AT4g22410" |
| | SHDVVWLKEYEDYKSWTKLYRPKPESVESVEFIRPLICSKDRSKILLEINNAALM | | exon | | complement(7775. .8034) |
| | WFDLESQSLTTAGIECDSSFTADILVSLVLCCKGDPDTQAQRSKDKQKMPKSTKRWA | | | | /number=6 |
| | LL" | | LTR | | 8059. .8449 |
| | complement(1789. .2973) | | | | /note="389bp LTR" |
| | /gene="AT4g22390" | | misc_feature | | 8450. .13234 |
| | /number=1 | | | | /note="LTR retrotransposon" |
| gene | 4728. .5711 | | | | 13234. .13624 |
| | /gene="AT4g22400" | | | | /note="389bp LTR" |
| CDS | 4728. .5711 | | | | 13513. .14331 |
| | /gene="AT4g22400" | | | | /gene="AT4g22420" |
| | /note="similarity to various predicted proteins, Arabidopsis thaliana | | | | complement(join(13513. .13618,13701. .14033,14216. .14331)) |
| | contains EST gb:Z17595" | | | | /gene="AT4g22420" |
| | /codon_start=1 | | | | complement(join(13513. .13618,13701. .14033,14216. .14331)) |
| | /product="putative protein" | | | | /gene="AT4g22420" |
| | /protein_id="CAB52814.1" | | | | /note="similarity to predicted protein, Arabidopsis thaliana |
| | /db_xref="GI:5738371" | | | | contains EST gb:T44427" |
| | /translation="MLSSHLSQICFSCRDPLGNLPADLLIKCTDMLNPAEGORLRT | | | | /codon_start=1 |
| | LKNWKLALPSFRKRAYREERPMWLYRSGSGTREFDPVRRVHSGNDPRADARFL | | | | /product="putative protein" |
| | GSTLGLVMSLSDLNPKRTHOFFLYNPFISELQQLPELTLDSPYPTGLRVGLVGN | | | | /protein_id="CAB52816.1" |
| | PDNNYACLIIDYLHIEIGNRDTYLLIYYAKTSGRWEQNWVASMVLVGLFDTWY | | | | /db_xref="GI:5738373" |
| | IESISPSDKISIEIWPKRDFDTFQEWNVSNFIPVDIQWDPFDQSLYHVKKRLQ | | | | /translation="MAECTPTDLINEMFLRLRATTLKKCRVLSKPCFSLIDSPEKRVIE |
| | LPENPSTLAIIGTSDTRVRNKKIIGQSDDDPRASDHIIIDGVWVVEVRP" | | | | |
| | 4728. .5711 | | | | |
| exon | /gene="AT4g22400" | | | | |
| | /number=1 | | | | |
| gene | 6112. .8034 | | | | |

```

RSDSPPLPLGFNNHLLPLANAYDDDDDEEGNELKKSQARRNCGVAKGNGKNVGEAQ
EMVDEDDDDDDSKRGKHSRSHVRRDPCYLDTVNRQVILIDQLMLRVLATMRK
RRMTGGRAKAKEMKYLKVKKSG"
complement(13513. .13618)
/gene="AT4g22420"
/number=1
complement(13619. .13700)
/gene="AT4g22420"
/number=1
complement(13701. .14033)
/gene="AT4g22420"
/number=2

Query Match      8.2%; Score 39.4; DB 8; Length 106702;
Best Local Similarity 59.3%; Pred. No. 37;
Matches 67; Conservative 0; Mismatches 46; Indels 0; Gaps 0;

QY 85 GTTGGCGCGCCACCGTGGTGGCGGTGGTGGCGCCAGCTGTCGGTAAGGC 144
||||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 33426 GTTGGAGGTGTCGCCGTGGTGGTGGAGGAGTTCCTGGTGGTGTTCGAGTGC 33485
||||| | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 145 TCTGGTGGTGTGTCGATGCCATCGCGGTGGCGGTCTCTGGCGGTCTGATTGG 197
||||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 33486 TCTGGTGGTGTCTTTCGAGGAGTTCGCCGTGGTGTCTTTCGAGGTCTTCTGG 33538
||||| | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 14
LOCUS ATCHRIV57 199577 bp DNA linear PLN 16-MAR-2000
DEFINITION Arabidopsis thaliana DNA chromosome 4, contig fragment No. 57.
ACCESSION AL161557
VERSION AL161557.2 GI:7269071
KEYWORDS
SOURCE Arabidopsis thaliana.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; Core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
1 (bases 48809 to 64829)
Robben,J., Grymonprez,B., Volckaert,G., Mewes,H.W., Lemcke,K. and
Mayer,K.F.X.
Unpublished
REFERENCE 2 (bases 64578 to 155510)
Wedler,H., Wambutt,R., Mewes,H.W., Lemcke,K. and Mayer,K.F.X.
Unpublished
REFERENCE 3 (bases 141638 to 141797)
Volckaert,G., Grymonprez,B., Voet,M., Robben,J., Mewes,H.W.,
Lemcke,K. and Mayer,K.F.X.
Unpublished
REFERENCE 4 (bases 143670 to 199577)
Hilbert,H., Braun,M., Holzer,E., Brandt,A., Duesterhoeft,A.,
Mewes,H.W., Lemcke,K. and Mayer,K.F.X.
Unpublished
REFERENCE 5 (bases 1 to 199577)
EU Arabidopsis sequencing, project.
Direct Submission
JOURNAL Submitted (10-MAR-2000) MIPS, at the Max-Planck-Institut fuer
Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG, E-mail:
lemcke@mips.biochem.mpg.de,mayer@mips.biochem.mpg.de
Coordinator: Mike Bevan, Molecular Genetics Department, Cambridge
Laboratory, John Innes Centre, Colney Lane, NR4 7UJ Norwich, UK,
E-mail: michael.bevan@bbsrc.ac.uk
Information on performance of analysis and a more detailed
annotation of this entry and other sequences of chromosomes 3, 4
and 5 can be viewed at: http://www.mips.biochem.mpg.de/proj/thal/
this fragment has an overlap with ATCHRIV56 at the 5' end and an
overlap with ATCHRIV58 at the 3' end.

FEATURES
Location/Qualifiers
1..199577
/organism="Arabidopsis thaliana"
/variety="Columbia"
/db_xref="taxon:3702"
/chromosome="4"

6674. .9057
/gene="AT4g22260"
complement(join(6674. .6882,7268. .7339,7507. .7575,
7667. .7775,7863. .7978,8361. .8454,8588. .8777,8909. .9057))
/gene="AT4g22260"
complement(join(6674. .6882,7268. .7339,7507. .7575,
7667. .7775,7863. .7978,8361. .8454,8588. .8777,8909. .9057))
/gene="AT4g22260"
/note="similarity to alternative oxidase, Mangifera
indica, PIR2:S45035
contains EST gb:T42793, AI994896.1, AI997645.1, R30022,
AA395166"
/codon_start=1
/product="putative protein"
/protein_id="CAB79181.1"
/db_xref="GI:7269072"
/translation="MAALSGISSGTLTISRPLVILRRRAAVSVSSSRLLHLPLSS
RLLLNHRVOATILQDDEKVVVEESFKATSTGTPEEPNNSSSTSAFETWII
KLEQGVNVELDSVIKILDTLYRDTYARFVLETIARVPYFAFMSVLMYETFCGWR
RADLYKHFAESWNEHMLIMEELGNSWDFRFLAQHIAITFYFMTVFLYILSPRM
AYHSECVESHAYETYDKFLKASGEELKNMPADIAVKYTGGLYLFENLYDFVNI
RDEAEHCCKTRACQTLGSLRSPHSILEDDEEESGCVPEEAHCEIYDCLKKSIT
S"
complement(6674. .6882)
/gene="AT4g22260"
/number=1
complement(6883. .7267)
/gene="AT4g22260"
/number=1
complement(7268. .7339)
/gene="AT4g22260"
/number=2
complement(7340. .7506)
/gene="AT4g22260"
/number=2
complement(7507. .7575)
/gene="AT4g22260"
/number=3
complement(7576. .7666)
/gene="AT4g22260"
/number=3
complement(7667. .7775)
/gene="AT4g22260"
/number=4
complement(7776. .7862)
/gene="AT4g22260"
/number=4
complement(7863. .7978)
/gene="AT4g22260"
/number=5
complement(7979. .8360)
/gene="AT4g22260"
/number=5
complement(8361. .8454)
/gene="AT4g22260"
/number=6
complement(8455. .8587)
/gene="AT4g22260"
/number=6
complement(8588. .8777)
/gene="AT4g22260"
/number=7
complement(8778. .8908)
/gene="AT4g22260"
/number=7
complement(8909. .9057)
/gene="AT4g22260"
/number=8
10103. .12393
/gene="AT4g22270"
join(10103. .10244,10395. .10516,11327. .11779,11868. .12174,
12254. .12393)
/gene="AT4g22270"
```

[illegible]

Search completed: February 22, 2003, 16:17:27
Job time : 1851.95 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 22, 2003, 14:26:45 ; Search time 963.886 Seconds
(without alignments)
8165.910 Million cell updates/sec

Title: US-09-677-374-1
Perfect score: 486
Sequence: 1 atgaacagaggtgttgca.....aagctattcaacagaaaa 486

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : EST.*

```

1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rod:*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----------|--------------------|
| 1 | 53.8 | 11.1 | 998 | BH133989 | BH133989 ENTNM17TF |
| C 2 | 52.4 | 10.8 | 868 | BH162606 | BH162606 ENTRV08TF |
| C 3 | 49.6 | 10.2 | 279 | CNS03W9F | AL263436 Tetraodon |
| C 4 | 47.2 | 9.7 | 845 | AZ685166 | AZ685166 ENTLG11TF |
| C 5 | 46.4 | 9.5 | 325 | CNS03PK8 | AL284753 Tetraodon |
| C 6 | 46.4 | 9.5 | 624 | BJ001798 | BJ001798 BJ001798 |

ALIGNMENTS

```

RESULT 1
BH133989
LOCUS ENTNM17TF Entamoeba histolytica Sheared DNA Entamoeba histolytica
DEFINITION ENTNM17TF Entamoeba histolytica Sheared DNA Entamoeba histolytica
ACCESSION BH133989
VERSION BH133989.1 GI:15093050
KEYWORDS GSS.
SOURCE Entamoeba histolytica.
ORGANISM Entamoeba histolytica.
REFERENCE Entamoeba histolytica.
AUTHORS Eukaryota; Entamoebidae; Entamoeba.
TITLE 1 (bases 1 to 998)
JOURNAL Loftus,B., Wang,Z., Van Aken,S. and Fraser,C.
COMMENT Determination of clone end sequences from Entamoeba histolytica
HM1:IMSS sheared DNA library (2001)
Unpublished (2001)
Contact: Brendan J Loftus
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0208
Fax: 301 838 3543
Email: bjl@tustigr.Org
Clones are derived from the Entamoeba histolytica HM1:IMSS sheared
DNA library
Seq primer: M13-Forward
Class: shotgun
High quality sequence start: 17
High quality sequence stop: 490.

```

```

c 7 46.4 9.5 872 13 BJ009355
c 8 46 9.5 909 17 CNS000JTL
c 9 46 9.5 939 17 CNS000CNG
c 10 44.6 9.2 890 17 BH164851
c 11 43.2 8.9 427 9 AU052962
c 12 43.2 8.9 440 9 AU052941
c 13 42.8 8.9 715 13 BJ406237
c 14 42.8 8.8 630 9 AL677875
c 15 42.6 8.8 442 12 BE801262
c 16 42.6 8.8 452 13 BI471035
c 17 42.6 8.8 536 17 CNS0300K
c 18 42.2 8.7 748 13 BI176637
c 19 42.2 8.7 805 17 AZ753057
c 20 41.8 8.6 562 13 BJ362836
c 21 41.8 8.6 723 17 AZ627893
c 22 41.8 8.6 745 17 CNS01MHE
c 23 41.6 8.6 294 9 AU060898
c 24 41.2 8.5 1101 17 CNS001L00
c 25 41 8.4 328 9 AI437626
c 26 41 8.4 344 12 BG352395
c 27 41 8.4 353 9 AI443457
c 28 41 8.4 402 9 AI442486
c 29 41 8.4 407 14 BQ742064
c 30 41 8.4 431 10 BE555999
c 31 41 8.4 460 14 BQ299690
c 32 41 8.4 465 13 BM524492
c 33 41 8.4 467 14 BQ080883
c 34 41 8.4 485 13 BM524782
c 35 41 8.4 496 12 BE805704
c 36 41 8.4 509 10 AW734942
c 37 41 8.4 511 13 BM524641
c 38 41 8.4 520 13 BM093415
c 39 41 8.4 521 12 BG790034
c 40 41 8.4 536 13 BI784659
c 41 41 8.4 543 13 BI701733
c 42 41 8.4 569 14 BQ629443
c 43 41 8.4 595 14 BQ473338
c 44 40.8 8.4 805 17 CNS04RW2
c 45 40.6 8.4 551 14 BM959137

```


| | | | | | |
|-----------------------|--------------|-------|------------------|----------|------------|
| BASE COUNT | 324 a | 161 c | 95 g | 265 t | |
| ORIGIN | | | | | |
| Query Match | | 9.7% | Score 47.2 | DB 17 | Length 845 |
| Best Local Similarity | | 52.6% | pred. No. 0.0029 | | |
| Matches 103 | Conservative | | | | |
| | | 0 | Mismatches 93 | Indels 0 | Gaps 0 |

| | Best Local Similarity | Matches | Conservative | Mismatches | Indels | Gaps |
|----|-----------------------|---|--------------|------------|--------|------|
| QY | 74 | TTAGTGTCTCAAGAAAGTCGGAGCTCGCACTGGGGCTGTGTTGGCGGTGTTGCTGCCACG | 133 | | | |
| Db | 743 | TTGTCCTGTGTCATTTGGTGGTCTTCTGTTGGTGTGTTGATGGTGCCTGTTGGTGGTGC | 684 | | | |
| QY | 134 | TGTTTGGTAAAGGTAGTCGGTGCAGTTCGATGGCCATGGTGGGTCGTGTTTGGGTGGAT | 193 | | | |
| Db | 683 | TTGCTGTGCTGTTTGGTGGTGCATCTTGGTGGTGTGTTGGTGGTGCCTCATTTGGTGGT | 624 | | | |
| QY | 194 | TAAATTGTTCTTAAAAATCGGTCAATCGATGGATCAGCAGGATAAAATAAAGCTAAACAGA | 253 | | | |
| Db | 623 | TTGTTGGTGGTGTCTGTTGGTGTAGTCATGTTATGAAAAAGACAAGTAACACCAAAAAA | 564 | | | |
| QY | 254 | GTTTGGAAAAAGGTAAA | 269 | | | |
| Db | 563 | TTGAAAAATTAAATAAA | 548 | | | |

| | | | | | |
|------------|---|------------|-----|--------|-----------------|
| RESULT 5 | CNS03PK8/c | 325 bp | DNA | linear | GSS 17-MAY-2000 |
| LOCUS | | | | | |
| DEFINITION | Tetraodon nigroviridis genome survey sequence PUC-Ori end of clone 045118 of library G from Tetraodon nigroviridis, genomic survey sequence. | | | | |
| ACCESSION | AL254753 | | | | |
| VERSION | AL254753.1 | GI:7975765 | | | |
| KEYWORDS | GSS; genome survey sequence. | | | | |
| SOURCE | Tetraodon nigroviridis. | | | | |
| ORGANISM | Tetraodon nigroviridis | | | | |
| | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes; Tetraodontidae; Tetraodon. | | | | |
| REFERENCE | 1 (bases 1 to 325) | | | | |
| AUTHORS | Roest-Crollius, H., Jaillon, O., Dasilva, C., Bouneau, L., Fisher, C., Bernot, A., Fizames, C., Wincker, P., Brottier, P., Quetier, F., Saurin, W. and Weissenbach, J. | | | | |
| | Human gene number estimate provided by genome wide analysis using Tetraodon nigroviridis DNA sequence | | | | |
| TITLE | Unpublished | | | | |
| JOURNAL | 2 (bases 1 to 325) | | | | |
| REFERENCE | | | | | |

| | |
|-----------|--|
| REFERENCE | 3 (bases 1 to 325) |
| AUTHORS | Genoscope. |
| TITLE | Direct Submission |
| JOURNAL | Submitted (12-APR-2000) |
| COMMENT | This sequence is a single read and was generated as part of a large scale clone-end sequencing project of the Tetraodon nigroviridis genome. For more information, please take a look at http://www.genoscope.cns.fr/Tetraodon . |
| FEATURES | Location/Qualifiers |
| source | 1..325 /organism="Tetraodon nigroviridis" |

```
/db_xref="taxon:99883"
/clone="045118"
/clone_lib="G"
/notes="genoscope sequence ID : C0B045BE09SPl-end :
PUC-Orl"
BASE COUNT 157 a 157 c 10 g 0 t 1 others
ORIGIN

Query Match 9.5%; Score 46.4; DB 17; Length 325;
Best Local Similarity 52.6%; Pred. No. 0.0034;
Matches 101; Conservative 0; Mismatches 91; Indels 0; Gaps 0;

QY 22 GGTAGTACTTAAATATATACAGTGTGTTTATGTTGGTGTGCCAGAACATTTAGTCTG 81
Db 237 GGTGGTGGTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 178

QY 82 CAAGAAGTCGGAGCTCGGACTGGGCTGTTGTTGGCGGTGTTGTTGGCGGCTGTTGTTGTT 141
Db 177 GTTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 118

QY 142 AAAGGTAGTGGTGCAGTTCGAATGGCCATTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 201
Db 117 GCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 58

QY 202 TCTAAATCGGT 213
Db 57 GTTGTGTTGGT 46

RESULT 6
BJ001798/c 624 bp mRNA linear EST 05-DEC-2001
LOCUS BJ001798 MF01SSA cDNA Oryzias latipes cDNA clone MF01SSA027D02 5',
DEFINITION mRNA sequence.
ACCESSION BJ001798
VERSION BJ001798.1 GI:17364689
KEYWORDS EST.
SOURCE Japanese medaka.
ORGANISM Oryzias latipes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Percormorpha; Atherinomorpha;
Beloniformes; Adrianichthyidae; Oryziinae; Oryzias.
1 (bases 1 to 624)
Kohara,Y., Shin-I., Kimura,T., Narita,T., Jindo,T. and Takeda,H.
Medaka EST Project in Takeda's lab
Unpublished (2001)
Contact: Tadasu Shin-i
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshini@genes.nig.ac.jp.
Location/Qualifiers
1..624
/organism="Oryzias latipes"
/strain="Hd-rR"
/db_xref="taxon:8090"
/clone="MF01SSA027D02"
/clone_lib="MF01SSA cDNA"
/sex="mixture of female and male"
/tissue_type="whole embryo"
/dev_stage="segmentation stage 20 - 25"

BASE COUNT 208 a 248 c 64 g 104 t
ORIGIN

Query Match 9.5%; Score 46.4; DB 13; Length 624;
Best Local Similarity 52.0%; Pred. No. 0.0045;
Matches 104; Conservative 0; Mismatches 96; Indels 0; Gaps 0;

QY 14 GTTTCAGGTAGTACTCTAATATATATATATATATATATATATATATATATATATATATATAT 73
Db 762 GTTTCAGGTATTTGGTGGTACTGTAGACGATGGTTTGGTGGTGGTGGTGGTGGTGGTGGT 703

QY 74 TTAGTCGTCAAGAAGTCGGAGCTGGGCTGGGCTGGGCTGGGCTGGGCTGGGCTGGGCTGGG 133
Db 702 TTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 643

QY 134 TGTTTGGTAAAGGTAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 193
Db 642 TAGACGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 583

QY 194 TAATTGGTTCATAAATCGGT 213
Db 582 CTGGTGGTACTATAGACGGT 563
```

[illegible]

TITLE Determination of clone end sequences from Entamoeba histolytica
JOURNAL HMI:IMSS sheared DNA library (2001)
COMMENT Unpublished (2001)
Contact: Brendan J Loftus
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0208
Fax: 301 838 3543
Email: b.loftus@tigr.org
Clones are derived from the Entamoeba histolytica HMI:IMSS sheared
DNA library
Seq primer: M13-Forward
Class: shotgun
High quality sequence start: 15
High quality sequence stop: 633.
Location/Qualifiers

```

1..890
/oranism="Entamoeba histolytica"
/train="HMI:IMSS"
/db_xref="taxon:5759"
/cloned_lib="Entamoeba histolytica Sheared DNA"
/note=Vector; PHOS1; Site_1: Bst I; Constructed at The
Institute for Genomic Research (TIGR), Rockville, MD.
Genomic DNA isolated from broth cultures of E. histolytica
using a method described by Clark and Diamond (Clark,
C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a
method for isolate identification. Exp. Parasitol.
77:450.). The DNA was mechanically sheared to give a
tight size distribution (~2 kb). The v + i method used for
the library construction is described in detail in Smith,
H.O. and Venter, J.C. (Making small insert libraries for
whole genome shotgun sequencing projects. In Genome
Sequencing: A Practical Approach, eds. M. Vaudin and B.
Barell, Oxford University Press, 1999)."
BASE COUNT      334 a    65 c   212 g   279 t
ORIGIN
Query Match          9.2%; Score 44.6; DB 17; Length 890;
Best Local Similarity 65.7%; Pred. No. 0.019;
Matches 65; Conservative 0; Mismatches 34; Indels 0; Gaps 0;

QY 103 GGCGCTCTGTGGCGTGTTGTCGCCAGCTGTTTGGTAAGTAGTGTGCAGATTGCA 162
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 596 GTGGCTGTGTGGTGGTGTATGCTGTGGTGGTGCTGCTGGTGGTGGTGGTGGT 655
|| |||||||||||||||| |||| ||||||||

QY 163 ATGCCAATTGGTGTGCTGTTTGGTGGATAATTGGT 201
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 656 GGTGCTGTGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 694

RESULT 11
LOCUS AU052962                               427 bp     mRNA       linear EST 28-APR-1999
DEFINITION AU052962 Dictyostelium discoideum SL (H.Urushi-hara) Dictyostelium
discoideum CDNA clone SFLF389, mRNA sequence.
ACCESSION AU052962
VERSION AU052962.1 GI:4701445
KEYWORDS EST.
SOURCE Dictyostelium discoideum.
ORGANISM Dictyostelium discoideum.
REFERENCE Eukaryota; Mycetozoa; Dictyostellida; Dictyostelium.
AUTHORS Morio,T., Urushihara,H., Saito,T., Ugawa,Y., Mizuno,H., Yoshida,M.,
Williams,J., Maeda,B.N., Pi,M., Sato,T., Takemoto,K., Yasukawa,H.,
Developmental cdna in Dictyostelium discoideum
Unpublished (1998)
CONTACT: Hideko Urushihara
Institute of Biological Sciences
University of Tsukuba
1-1-1 Tennoudai, Tsukuba, Ibaraki 305-8572, Japan
TEL: 81-298-53-4664
```

[illegible]

GenCore version 5.1.4_p5-4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 22, 2003, 14:36:28 ; Search time 40.2302 Seconds
(without alignments)
3699.288 Million cell updates/sec

Title: US-09-677-374-1

Perfect score: 486

Sequence: 1 atgacacagaggtgttgc.....aagtcatttcacacagaaaaa 486

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents.NA.*
1: /cgn2.6/ptodata/1/ina/5A_COMB.seq.*
2: /cgn2.6/ptodata/1/ina/5b_COMB.seq.*
3: /cgn2.6/ptodata/1/ina/6A_COMB.seq.*
4: /cgn2.6/ptodata/1/ina/6B_COMB.seq.*
5: /cgn2.6/ptodata/1/ina/PTCUS_COMB.seq.*
6: /cgn2.6/ptodata/1/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|---------------------|
| C 1 | 47.8 | 9.8 | 7218 | 1 | US-08-232-463-14 |
| C 2 | 42.6 | 8.8 | 465 | 4 | US-08-476-1028-2 |
| C 3 | 35.8 | 7.4 | 303 | 4 | US-08-556-978B-80 |
| C 4 | 35.8 | 7.4 | 303 | 4 | US-08-556-978B-81 |
| C 5 | 35.4 | 7.3 | 303 | 4 | US-08-556-978B-82 |
| C 6 | 35.2 | 7.2 | 198 | 5 | PCT-US95-10668-3 |
| C 7 | 35.2 | 7.2 | 198 | 5 | PCT-US95-10668-4 |
| C 8 | 34.4 | 7.1 | 1046 | 1 | US-08-361-467B-4 |
| C 9 | 34.4 | 7.1 | 1046 | 1 | US-08-484-332C-4 |
| C 10 | 33.6 | 6.9 | 198 | 5 | PCT-US95-10668-1 |
| C 11 | 33.6 | 6.9 | 198 | 5 | PCT-US95-10668-2 |
| C 12 | 33.2 | 6.8 | 1011 | 2 | US-08-825-781-2 |
| C 13 | 33.2 | 6.8 | 2072 | 2 | US-09-073-362-2 |
| C 14 | 33.2 | 6.8 | 2072 | 2 | US-09-243-920-2 |
| C 15 | 33 | 6.8 | 377 | 2 | US-08-332-766A-1 |
| C 16 | 32.8 | 6.7 | 1690 | 1 | US-08-276-452A-24 |
| C 17 | 32.8 | 6.7 | 1690 | 2 | US-08-798-744-24 |
| C 18 | 32.4 | 6.7 | 1608 | 4 | US-09-292-225-20 |
| C 19 | 32.4 | 6.7 | 1608 | 4 | US-09-292-225-22 |
| C 20 | 32.4 | 6.7 | 1665 | 4 | US-09-292-225-17 |
| C 21 | 32.4 | 6.7 | 1665 | 4 | US-09-292-225-19 |
| C 22 | 32.4 | 6.7 | 1752 | 4 | US-09-292-225-14 |
| C 23 | 32.4 | 6.7 | 1752 | 4 | US-09-292-225-16 |
| C 24 | 32.4 | 6.7 | 1984 | 1 | US-07-885-970A-25 |
| C 25 | 32.4 | 6.7 | 1985 | 1 | US-08-298-687A-25 |
| C 26 | 32.4 | 6.7 | 1985 | 1 | US-08-298-829-25 |
| C 27 | 32.2 | 6.6 | 1440 | 4 | US-09-134-001C-1598 |

C 28 32.2 6.6 1561 1 US-07-968-971A-3
C 29 32.2 6.6 1561 1 US-07-824-247-44
C 30 32.2 6.6 1561 1 US-08-142-473A-4
C 31 32.2 6.6 1561 1 US-08-469-203A-4
C 32 32.2 6.6 1561 1 US-08-469-203A-4
C 33 32.2 6.6 1561 3 US-08-470-204A-44
C 34 32.2 6.6 1561 4 US-08-926-522-10
C 35 31.8 6.5 379 1 US-08-145-617-5
C 36 31.8 6.5 2793 1 US-08-209-747-1
C 37 31.8 6.5 2793 1 US-08-458-298-1
C 38 31.8 6.5 5496 4 US-09-462-284-1
C 39 31.6 6.5 1632 1 US-08-324-243-34
C 40 31.6 6.5 1632 1 US-08-532-390-34
C 41 31.6 6.5 1632 3 US-08-717-294-34
C 42 31.6 6.5 1632 5 PCT-US95-11511-34
C 43 31.6 6.5 3656 4 US-09-221-017B-786
C 44 31.4 6.5 925 3 US-08-858-003-1
C 45 31.4 6.5 925 3 US-09-078-166-1

ALIGNMENTS

RESULT 1
US-08-232-463-14/c
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; CLONE: ptzpgt-Fls
; US-08-232-463-14

Query Match 9.8% Score 47.8; DB 1; Length 7218;

| | | | | | |
|---|------|---|------|------------|------|
| Best Local Similarity 2.0%; Pred.No.3.le-05; | | | | | |
| Matches | 7; | Conservative | 206; | Mismatches | 138; |
| | | | | Indels | 0; |
| Gaps | | | | | 0; |
| | | | | | |
| QY | 136 | TTTTGTTAAAGTAGTGGTCGATGTTCATATGCCATTGGTGCTGTTTGGGTGCATTGA | 195 | | |
| | | : | | | |
| Db | 1442 | TITGGTACRRR | 1383 | | |
| | | | | | |
| QY | 196 | ATTGGTTCAAATCGGTCAATCGATGATCAGCAGGATAAATAAACCTAACCCAGNT | 255 | | |
| | | : | | | |
| Db | 1382 | RR | 1323 | | |
| | | | | | |
| QY | 256 | TTGAAAAGGTAAAGCAGGCGCAAGTGACACGTTGGCGTAATCCAGATACAGGCAATAGT | 315 | | |
| | | : | | | |
| Db | 1322 | RR | 1263 | | |
| | | | | | |
| QY | 316 | TATAGTGTTCAGCCAGTGGTACTTACCAGCGTTACAATAAGCAAGCGTCGCCAGCAA | 375 | | |
| | | : | | | |
| Db | 1262 | RR | 1203 | | |
| | | | | | |
| QY | 376 | TATTGTCGAGAATTCACGAAAAAGCGATGATTGAGGCGACAGCAAGAGATTACGGC | 435 | | |
| | | : | | | |
| Db | 1202 | RR | 1143 | | |
| | | | | | |
| QY | 436 | ACTGCATCGCGCACCGGATGCTGTGGCAAGTCATTCAACAGAAAAA | 486 | | |
| | | : | | | |
| Db | 1142 | RR | 1092 | | |
| | | | | | |
| RESULT 2 | | | | | |
| US-08-476-102A-2 | | | | | |
| ; Sequence 2, Application US/08476102A | | | | | |
| ; Patent No. 6355450 | | | | | |
| GENERAL INFORMATION: | | | | | |
| APPLICANT: Fleischmann, et al. | | | | | |
| TITLE OF INVENTION: Nucleotide Sequence of the Haemphilus influenzae Rd Genome, Fragments Thereof, and Uses Thereof | | | | | |
| Patent No. 6355450 | | | | | |
| | | | | | |
| NUMBER OF SEQUENCES: 9 | | | | | |
| CORRESPONDENCE ADDRESS: | | | | | |
| ADDRESSER: Human Genome Sciences, Inc. | | | | | |
| STREET: 9410 Key West Ave. | | | | | |
| CITY: Rockville | | | | | |
| STATE: MD | | | | | |
| COUNTRY: USA | | | | | |
| ZIP: 20850 | | | | | |
| COMPUTER READABLE FORM: | | | | | |
| MEDIUM TYPE: Floppy disk | | | | | |
| COMPUTER: IBM PC compatible | | | | | |
| OPERATING SYSTEM: PC-DOS/MS-DOS | | | | | |
| SOFTWARE: Patenlin Release #1.0, Version #1.30 | | | | | |
| CURRENT APPLICATION DATA: | | | | | |
| APPLICATION NUMBER: US/08/476,102A | | | | | |
| FILING DATE: 07-Jun-1995 | | | | | |
| CLASSIFICATION: <Unknown> | | | | | |
| ATTORNEY/AGENT INFORMATION: | | | | | |
| NAME: Marks, Michelle S. | | | | | |
| REGISTRATION NUMBER: 41,971 | | | | | |
| REFERENCE/DOCKET NUMBER: PB186P2 | | | | | |
| TELECOMMUNICATION INFORMATION: | | | | | |
| TELEPHONE: 301-309-8504 | | | | | |
| TELEX: 301-309-8439 | | | | | |
| INFORMATION FOR SEQ ID NO: 2: | | | | | |
| SEQUENCE CHARACTERISTICS: | | | | | |
| LENGTH: 465 base pairs | | | | | |
| TYPE: nucleic acid | | | | | |
| STRANDEDNESS: single | | | | | |
| TOPOLOGY: linear | | | | | |
| MOLECULE TYPE: DNA (genomic) | | | | | |
| SEQUENCE DESCRIPTION: SEQ ID NO: 2: | | | | | |
| US-08-476-102A-2 | | | | | |

US-08-556-978B-81
; Sequence 81, Application US/08556978B
; Patent No. 6268169
; GENERAL INFORMATION:
; APPLICANT: FARNSTOCK, STEPHEN F.
; TITLE OF INVENTION: NOVEL RECOMBINANTLY PRODUCED
; TITLE OF INVENTION: SPIDER SILK ANALOGS
; NUMBER OF SEQUENCES: 107
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: E. I. DU PONT DE NEMOURS AND COMPANY
; STREET: 1007 MARKET STREET
; CITY: WILMINTON
; STATE: DELAWARE
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 19898
; COMPUTER READABLE FORM:
; MEDIUM TYPE: DISKETTE, 3.50 INCH
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: MICROSOFT WINDOWS 95
; SOFTWARE: MICROSOFT WORD FOR WINDOWS 95
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/556,978B
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/077,600
; FILING DATE: JUNE 15, 1993
; ATTORNEY/AGENT INFORMATION:
; NAME: FLOYD, LINDA AXAMETHY
; REGISTRATION NUMBER: 33,692
; REFERENCE/DOCKET NUMBER: CR-9389-A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 302-892-8112
; TELEFAX: 302-773-0164
; INFORMATION FOR SEQ ID NO: 81:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 303 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-556-978B-81

Query Match 7.4%; Score 35.8; DB 4; Length 303;
Best Local Similarity 56.3%; Pred. No. 0.04;
Matches 67; Conservative 0; Mismatches 52; Indels 0; Gaps 0;
QY 84 AGAAGTCGGAGCTGCAGCTGGGCTGTGTTGGCGGTGTTGCTGGCCAGCTGTTGGTAA 143
Db 83 AGCGCTGGCGGAGCTGGTGGTCTGGTCAAGGCGGCTTGGCTCAAGGCGCGGTCA 142
QY 144 AGCTAGTGGTCGAGTTGCAATGGCAATGGTGTGTTGTTGGTGGATTAATTGGTT 202
Db 143 AGCGCTGGTGCAGCAGCAGCTGCCGCTGGCGGTGCAGGCCAAGGTGGATATGGTGCT 201

RESULT 5
US-08-556-978B-82
; Sequence 82, Application US/08556978B
; Patent No. 6268169
; GENERAL INFORMATION:
; APPLICANT: FARNSTOCK, STEPHEN F.
; TITLE OF INVENTION: NOVEL RECOMBINANTLY PRODUCED
; TITLE OF INVENTION: SPIDER SILK ANALOGS
; NUMBER OF SEQUENCES: 107
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: E. I. DU PONT DE NEMOURS AND COMPANY
; STREET: 1007 MARKET STREET
; CITY: WILMINTON
; STATE: DELAWARE
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 19898
; COMPUTER READABLE FORM:

; MEDIUM TYPE: DISKETTE, 3.50 INCH
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: MICROSOFT WINDOWS 95
; SOFTWARE: MICROSOFT WORD FOR WINDOWS 95
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/556,978B
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/077,600
; FILING DATE: JUNE 15, 1993
; ATTORNEY/AGENT INFORMATION:
; NAME: FLOYD, LINDA AXAMETHY
; REGISTRATION NUMBER: 33,692
; REFERENCE/DOCKET NUMBER: CR-9389-A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 302-892-8112
; TELEFAX: 302-773-0164
; INFORMATION FOR SEQ ID NO: 82:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 303 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-556-978B-82

Query Match 7.3%; Score 35.4; DB 4; Length 303;
Best Local Similarity 53.2%; Pred. No. 0.054;
Matches 75; Conservative 0; Mismatches 66; Indels 0; Gaps 0;
QY 62 GTGCCAGAACTTTAGTCGTCAAGAGTCGGAGCTGGGCTGTTGTTGGCGGTG 121
Db 68 GTGCCAGGCTGCAGGTGCTGCTGCTGGCGTGTGCTGTCAGGTGAGTGCAGGTGCTC 127
QY 122 TTGCTGGCCAGCTGTTGGTAAAGGTAGTGTGCTGAGTTGCAATGGCCATGTTGGTGTG 181
Db 128 TGGGATCTCAGGCGCAGGTCAAGGTGCTGTCAGCTGGCGGCGCAGCTGGTGGCGCG 187
QY 182 TTTTGGGTGATTAAATTGGTT 202
Db 188 GTCAAGTGGCTAGCGCGGTT 208

RESULT 6
PCT-US95-10668-3/c
; Sequence 3, Application PC/TUS9510668
; GENERAL INFORMATION:
; APPLICANT: James Eberwine
; TITLE OF INVENTION: A Method of Sequencing Proteins by
; TITLE OF INVENTION: Epitope Ordering and Protein
; TITLE OF INVENTION: Restriction Mapping
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Jane Massey Licata, Esq.
; STREET: 210 Lake Drive East, Suite 201
; CITY: Cherry Hill
; STATE: NJ
; COUNTRY: USA
; ZIP: 08002
; COMPUTER READABLE FORM:
; MEDIUM TYPE: DISKETTE, 3.5 INCH, 1.44 Mb STORAGE
; COMPUTER: IBM 486
; OPERATING SYSTEM: WINDOWS FOR WORKGROUPS
; SOFTWARE: WORDPERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/10668
; FILING DATE: Herewith
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/294,133
; FILING DATE: August 22, 1994
; ATTORNEY/AGENT INFORMATION:

[illegible]

INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 1046 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA to mRNA
IMMEDIATE SOURCE:
CLONE: 3C9
US-08-361-467B-4

Query Match 7.1%; Score 34.4; DB 1; Length 1046;
Best Local Similarity 52.9%; Pred. No. 0.22; Gaps 0;
Matches 74; Conservative 0; Mismatches 66; Indels 0; Gaps 0;
QY 84 AGAAGTCGGAGTCGCGACTGGGCTGTGTGGCGGTGTGCTGCGCCAGCTGTTTGCTAA 143
DB 329 AGGAGCAGGATAAGGTGATGTTGGTGTGTTAACTGGTGGCTAGCTGGTGG 270
QY 144 AGGTAGTGGTCGAGTTCGAATGGCCATTGGTGGTGTGTTGGTGGGATTAATTGTTTC 203
DB 269 TTTTGTAGTGGCTATAAGTGGGAGCCTTAGTGTGGTGTAAACCGGTGTTAACTGGTGG 210
QY 204 TAAATCGGTCAATCGATGG 223
DB 209 TTTGATGGGCTATAAGTGG 190

RESULT 9

US-08-484-332C-4/c
Sequence 4, Application US/08484332C

Patent No. 5767374

GENERAL INFORMATION:

APPLICANT: De Greef, Willy
APPLICANT: Van Emmelo, John
APPLICANT: De Oliveria, Dulce E.
APPLICANT: De Souza, Maria-Helena
APPLICANT: Van Montagu, Marc
TITLE OF INVENTION: PLANTS WITH MODIFIED FLOWERS, SEEDS OR
NUMBER OF SEQUENCES: 13
CORRESPONDENCE ADDRESS:
ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS
STREET: P.O. Box 1404
CITY: Alexandria
STATE: Virginia
COUNTRY: United States
ZIP: 22313-1404
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/484,332C
FILING DATE: 7-JUNE-1995
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/361,467
FILING DATE: 22-DEC-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/681,492
FILING DATE: 04-APR-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO PCT/EP90/01275
FILING DATE: 01-AUG-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: EP 89 402 224.3
FILING DATE: 04-AUG-1989
ATTORNEY/AGENT INFORMATION:
NAME: Schulman, Robert M.
REGISTRATION NUMBER: 31,196
REFERENCE/DOCKET NUMBER: 010830-093

TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 836-6620
TELEFAX: (703) 836-2021
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 1046 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA to mRNA
IMMEDIATE SOURCE:
CLONE: 3C9
US-08-484-332C-4

Query Match 7.1%; Score 34.4; DB 1; Length 1046;
Best Local Similarity 52.9%; Pred. No. 0.22; Gaps 0;
Matches 74; Conservative 0; Mismatches 66; Indels 0; Gaps 0;
QY 84 AGAAGTCGGAGTCGCGACTGGGCTGTGTGGCGGTGTGCTGCGCCAGCTGTTTGCTAA 143
DB 329 AGGAGCAGGATAAGGTGATGTTGGTGTGTTAACTGGTGGCTAGCTGGTGG 270
QY 144 AGGTAGTGGTCGAGTTCGAATGGCCATTGGTGGTGTGTTGGTGGGATTAATTGTTTC 203
DB 269 TTTTGTAGTGGCTATAAGTGGGAGCCTTAGTGTGGTGTAAACCGGTGTTAACTGGTGG 210
QY 204 TAAATCGGTCAATCGATGG 223
DB 209 TTTGATGGGCTATAAGTGG 190

RESULT 10

PCT-US95-10668-1/c
Sequence 1, Application PC/TUS9510668
GENERAL INFORMATION:
APPLICANT: James Eberwine
TITLE OF INVENTION: A Method of Sequencing Proteins by
TITLE OF INVENTION: Epitope Ordering and Protein
RESTRICTION Mapping
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Jane Massey Licata, Esq.
STREET: 210 Lake Drive East, Suite 201
CITY: Cherry Hill
STATE: NJ
COUNTRY: USA
ZIP: 08002
COMPUTER READABLE FORM:
MEDIUM TYPE: DISKETTE, 3.5 INCH, 1.44 Mb STORAGE
COMPUTER: IBM 486
OPERATING SYSTEM: WINDOWS FOR WORKGROUPS
SOFTWARE: WORDPERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/10668
FILING DATE: Herewith
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/294,133
FILING DATE: August 22, 1994
ATTORNEY/AGENT INFORMATION:
NAME: Jane Massey Licata
REGISTRATION NUMBER: 32,257
REFERENCE/DOCKET NUMBER: PENN-0137
TELECOMMUNICATION INFORMATION:
TELEPHONE: (609) 779-2400
TELEFAX: (609) 779-8488
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 198
TYPE: Nucleic Acid
STRANDEDNESS: Single
TOPOLOGY: Linear
ANTI-SENSE: No

Search completed: February 22, 2003, 17:07:09
Job time : 63.2902 secs

QY 202 TCTAAATCGGT 213
Db 45757 GGTGATGGTGGT 45768

RESULT 2

US-09-864-761-19900
; Sequence 19900, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Acomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 19900
; LENGTH: 522
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AL049749.1
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 5.8
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 4.2
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 5.8
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 3.2
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 4.7
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 5.7
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 5.2
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 5.3
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 3.9

OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 4.4
OTHER INFORMATION: EST_HUMAN HIT: AA776664.1, EVALUE 4.90e-01
US-09-864-761-19900

Query Match 8.1%; Score 39.6; DB 10; Length 522;
Best Local Similarity 53.2%; Pred No. 0.014; 74; Indels 0; Gaps 0;
Matches 84; Conservative 0; Mismatches 74; Indels 0; Gaps 0;

QY 33 AATTATTATCAGTGTGTTTTTGTGGTGTGCCAGAACTTTAGTCGTCAGAAAGTCGG 92
Db 121 AGTGATCATGAAGTGTGATGATGTTGGTGTGTAATAATGACTGTGATAAAGATGGTGG 180
QY 93 ACCTGGGACTGGGGCTGTTGGGGCTGTGCTGCCACAGCTGTTGGTAAAGAGTGTAGTGG 152
Db 181 TGATGAAGCTGTGATGTTGGTGTGATGTTGGTGTGATGATGATGATGATGATGATGATG 240
QY 153 TCGAGTGTCAATGGCCATGTTGGTGTCTCTTTTGGGTG 190
Db 241 TGTAGTGTGATGTTGGCAGTGAATGTTGGTGTGTAATGGTG 278

RESULT 3

US-09-864-761-23424
; Sequence 23424, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO
; FILE REFERENCE: Acomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1


```
; SEQ ID NO 23424
; LENGTH: 350
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC019159.3
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 5.7
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 5.1
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 5.3
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 4.5
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 4.6
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 3.6
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 4.2
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 5.4
; OTHER INFORMATION: NT HIT: AF111944.1, EVALUAE 4.00e-03
; OTHER INFORMATION: EST_HUMAN HIT: AA484400.1, EVALUAE 3.20e-01
US-09-864-761-23424

Query Match      8.0%; Score 39; DB 10; Length 350;
Best Local Similarity 50.8%; Pred. No. 0.018;
Matches 93; Conservative 0; Mismatches 90; Indels 0; Gaps 0;

QY 7 AGAGGATGTTTGCAGGCTAGTCTAATATTATTCAGTGTGTTTATTAGTGGCTGTGCC 66
    || || || || || || || || || || || || || || || || || || || ||
Db 52 AGTGGTGATTGATGCTGTGGTGGCGCATATCTTAATGCTGATGCTGATGGTGCT 111
QY 67 CAGAACTTTAGTCGTCGAAGAGTCGGAGCTGGGACTGGGCTGTGTTTGGCGGTGTGCT 126
    || || || || || || || || || || || || || || || || || || || ||
Db 112 GATGGGGGTGGTGTGATGGAGGTGATGATGATGGTGTGCTGGTGTGATGATGAT 171
QY 127 GGCAGCTGTTTGGTAAAGCTAGTGTGTCGAGTTCGAATGCCATTGCTGCTGTTTGG 186
    || || || || || || || || || || || || || || || || || || || ||
Db 172 GGTGGTGGTGGCGCAATGTTGGTGATGATGATGATGATGATGATGATGATGATGAT 231
QY 187 GGT 189
    |||
Db 232 GGT 234

RESULT 4
US-09-864-761-6690
; Sequence 6690, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Hanzel, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aeomica-X-1
; CURRENT FILING DATE: 2001-05-23
; PRIOR FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
```

```
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 6690
; LENGTH: 600
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC019159.3
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 5.7
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 5.1
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 5.3
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 4.5
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 4.6
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 3.6
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 4.2
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 5.4
US-09-864-761-6690

Query Match      8.0%; Score 39; DB 10; Length 600;
Best Local Similarity 50.8%; Pred. No. 0.024;
Matches 93; Conservative 0; Mismatches 90; Indels 0; Gaps 0;

QY 7 AGAGGATGTTTGCAGGCTAGTCTAATATTATTCAGTGTGTTTATTAGTGGCTGTGCC 66
    || || || || || || || || || || || || || || || || || || || ||
Db 216 AGTGGTGATTGATGCTGTGGTGGCGCATATCTTAATGCTGATGCTGATGGTGCT 275
QY 67 CAGAACTTTAGTCGTCGAAGAGTCGGAGCTGGGACTGGGCTGTGTTTGGCGGTGTGCT 126
    || || || || || || || || || || || || || || || || || || || ||
Db 276 GATGGGGGTGGTGTGATGGAGGTGATGATGATGGTGTGCTGGTGTGATGATGAT 335
QY 127 GGCAGCTGTTTGGTAAAGCTAGTGTGTCGAGTTCGAATGCCATTGCTGCTGTTTGG 186
    || || || || || || || || || || || || || || || || || || || ||
Db 336 GGTGGTGGTGGCGCAATGTTGGTGATGATGATGATGATGATGATGATGATGATGAT 395
QY 187 GGT 189
    |||
Db 396 GGT 398

RESULT 5
US-09-864-761-21484
; Sequence 21484, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aeomica-X-1
; CURRENT FILING DATE: 2001-05-23
; PRIOR FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
```

```

RESULT 6
US-09-864-761-4745
; Sequence 4745, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aeomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annonax Sequence Listing Engine vers. 1.1
; SEQ ID NO 4745
; LENGTH: 1981
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC005414.2
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 5.5
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 6.2
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 7.1
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 6.1
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 5.5
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 8.2
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 7.7
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 3.2
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 7.9
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 11
US-09-864-761-4745

```

; NUMBER OF SEQ ID NOS: 43117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1

```
Best Local Similarity 50.0%; Pred. No. 0.052;
Matches 97; Conservative 0; Mismatches 97; Indels 0; Gaps 0;

QY 8 GAGGATGTTTGCAGGTAGTGTCTAATATTATACAGTGTGTTTTAGTTGGCTGTGCC 67
    |||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 291 GAGGTGTTGTTGGTGGCAATGATGTAATGTAAGTGTGTTGGATGGTGAATG 350

QY 68 AGAATTCTAGTCTCAAGAGTGGAGCTGCGACTGGGGCTGTGTTGGCGGTGTTGCTG 127
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 351 GGGATGTTGTAATGCTGATGAGTGTGATGAGAGTGTGATGTTGGTGGGGATG 410

QY 128 GCCAGCTGTTTGTAAAGTAGTGGTGGAGCTTGCAATGGCCATGTTGGTGTCTTTGG 187
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 411 GTGATGATGTTGGTGGTAATGGTGTGTAATGATGTTGGTGTGATGATGGTGGTGTG 470

QY 188 GTGGATTAAATGGT 201
    |||| | | | |
Db 471 GTGGTGAATGGT 484

RESULT 7
US-09-815-242-6926
; Sequence 6926, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011a
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6926
; LENGTH: 1671
; TYPE: DNA
; ORGANISM: Haemophilus influenzae
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1671)
US-09-815-242-6926

Query Match 7.9%; Score 38.4; DB 10; Length 1671;
Best Local Similarity 55.1%; Pred. No. 0.063;
Matches 75; Conservative 0; Mismatches 61; Indels 0; Gaps 0;

QY 18 GCAAGGTAGTGTCTAATATTATACAGTGTGTTTTAGTTGGTGTGCCAGAACTTTAG 77
    |||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 396 GCAAGGTTGCAATGTAAGTGAACCCCGTGGTCAAGTTGGGCTGGCAATGAATAC 455

QY 78 TCGTCAAGAAGTGGGAGCTGGAGTGGGGCTGTGTTGGCGGTGTTCTGCTGCCAGCTGTT 137
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
```

```
Db 456 TGTGGAAGAAGTCGCACCGCTGATTGGTATTGTTGCACGAGATATTGACGTGCCTTT 515

QY 138 TGTAAAGTAGTGGT 153
    | | | | | | | | | |
Db 516 AGATAAATTTAAAGT 531

RESULT 8
US-09-864-761-20699
; Sequence 20699, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Hanzel, David R.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO
; FILE REFERENCE: Aeomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 20699
; LENGTH: 446
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC007249.2
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 10
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 14
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 13
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 12
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 13
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 13
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 12
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 13
```

```

RESULT 10
US-09-864-761-2513
; Sequence 2513, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEOTIDE SEQUENCES
; FILE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Aeonica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668

```


FILE REFERENCE: Aomica-X-1
CURRENT APPLICATION NUMBER: US/09/864,761
PRIOR FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263.6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
SEQ ID NO 16846
LENGTH: 249
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AC007263.3
OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 8.3
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 7
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 8.1
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 7.5
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 8.8
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 9.5
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 9.2
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 7.6
OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 6.1
OTHER INFORMATION: NT HIT: AF164623.1, EVALUE 5.10e-01
OTHER INFORMATION: EST_HUMAN HIT: AW846120.1, EVALUE 1.00e-02
OTHER INFORMATION: SWISSPROT HIT: Q12329, EVALUE 4.80e+00

Query Match 7.7%; Score 37.4; DB 10; Length 249;
Best Local Similarity 55.9%; Pred. No. 0.046;
Matches 71; Conservative 0; Mismatches 56; Indels 0; Gaps 0;
QY 75 TAGTCGTCAAGAGTCGGAGCTGCGACTGGGCTGTGTGGCGGTGTGGCCAGCT 134
DB 10 TGGTGTAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 69
QY 135 GTTGTGTAAGAGTAGTGTGCGAGTTGCAATGGCCATTGGTGTGTTTGGTGGATT 194
DB 70 TGGTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 129

QY 195 AATTGGT 201
DB 130 TCGTGGT 136
RESULT 13
US-09-864-761-6
; Sequence 6, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO
; FILE REFERENCE: Aomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 6
; LENGTH: 476
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC007263.3
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 8.3
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 7
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 8.1
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 7.5
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 8.8
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 9.5
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 9.2
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 7.6
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 6.1

GenCore version 5.1.4_p5-4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 22, 2003, 13:08:35 ; Search time 135.699 Seconds
(without alignments)
8065.404 Million cell updates/sec

Title: US-09-677-374-1

Perfect score: 486

Sequence: 1 atgaacagaggtgttgca.....aagtcatttcacagaaaaa 486

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N_Geneseq_101002.*
1: /SID22/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
2: /SID22/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
3: /SID22/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
4: /SID22/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
5: /SID22/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
6: /SID22/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
7: /SID22/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
8: /SID22/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
9: /SID22/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
10: /SID22/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
11: /SID22/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
12: /SID22/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
13: /SID22/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
14: /SID22/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
15: /SID22/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
16: /SID22/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
17: /SID22/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
18: /SID22/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
19: /SID22/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
20: /SID22/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
21: /SID22/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
22: /SID22/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
23: /SID22/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
24: /SID22/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|-------------|
| 1 | 486 | 100.0 | 486 | 22 | AAF86246 |
| 2 | 486 | 100.0 | 486 | 24 | ABK52401 |
| 3 | 486 | 100.0 | 489 | 22 | AAH79040 |
| 4 | 295 | 60.7 | 483 | 22 | AAF86247 |
| 5 | 295 | 60.7 | 483 | 24 | ABK52402 |
| 6 | 295 | 60.7 | 768 | 22 | AAF86248 |
| 7 | 295 | 60.7 | 768 | 24 | ABK52403 |
| 8 | 73.2 | 15.1 | 110 | 22 | AAF86252 |
| 9 | 73.2 | 15.1 | 110 | 24 | ABK52407 |

PCR primer #4 used
Escherichia coli c
PCR primer #5 used
Escherichia coli c
PCR primer #3 used
Escherichia coli c
Rickettsia 17 kD o
Escherichia coli c
Drosophila melanog
Human genomic regi
H.Influenzae DNA f
Sequence of Haemop
DNA of the PAC clo
Human PAC clone se
Drosophila melanog
Drosophila melanog
Listeria monocytog
DNA encoding novel
Arabidopsis thalia
Listeria monocytog
Human breast cell
Human foetal liver
Probe #13046 for g
Human brain expres
Human bone marrow
Probe #12357 for g
Probe #16401 used
Probe #8101 used t
Human genome-deriv
DNA encoding novel
Human foetal liver
Probe #16570 for g
Human brain expres
Human bone marrow

ALIGNMENTS

RESULT 1

AAF86246
ID AAF86246 standard; DNA; 486 BP.

XX
AC AAF86246;

DT 11-JUL-2001 (first entry)
XX OspA 17kD antigen gene.

DE
XX Poikilothermic fish; Piscirickettsia salmonis; rickettsial pathogen;
KW vaccine; OspA; salmonid rickettsial septicaemia; rickettsial disease;
KW SRS; ds.

XX Piscirickettsia salmonis.
OS
XX Key Location/Qualifiers
FH 1..486
CDS /*tag= a
FT /partial
FT /product= "OspA"
FT /note= "Genus specific 17kDa antigen, the sequence does not include a stop codon"

XX CA2281913-AL.
PN
XX 17-MAR-2001.
PD
XX 17-SEP-1999;
PF 99CA-2281913.
XX
PR 17-SEP-1999;
XX 99CA-2281913.

PA (KAYW/) KAY W W.
PA (BURL/) BURIAN J.
PA (KUZU/) KUZUK M A.

XX Kay WW, Burian J, Kuzyk MA;

XX WPI; 2001-316844/34.

XX P-PSDB; AAB81126.

XX Method for protecting poikilothermic fish against salmonid rickettsial
PT septicemia and other rickettsial diseases comprises administering a
PT vaccine containing the Ospa protein of *Piscirickettsia salmonis*

XX Disclosure; Fig 2B; 35pp; English.

XX This invention relates to a method for the protection against infection
CC of a poikilothermic fish by the bacterial pathogen, *Piscirickettsia*
CC salmonis. The method comprises administering an immunogenic amount of a
CC *P. salmonis* specific antigen termed Ospa, or an immunogenic fragment of
CC Ospa in the form of a vaccine. The method is used for protecting animals,
CC particularly poikilothermic fish, against the bacterial pathogen
CC *P. salmonis*. The method is also useful for protecting against salmonid
CC rickettsial septicemia (SRS) and other rickettsial diseases. The present
CC sequence represents DNA encoding the *P. salmonis* Ospa protein. An Ospa
CC protein with an N-terminal fusion partner is used in a vaccine to create
CC an anti-Ospa antibody response.

XX Sequence 486 BP; 137 A; 79 C; 144 G; 126 T; 0 other;

Query Match 100.0%; Score 486; DB 22; Length 486;
Best Local Similarity 100.0%; Pred. No. 4e-137;
Matches 486; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGAACAGAGGATGTTGCAAGTAGTCTAATATTATTATCAGTGTGTTAGTTGGC 60

Db 1 ATGAACAGAGGATGTTGCAAGTAGTCTAATATTATTATCAGTGTGTTAGTTGGC 60

QY 61 TGTGCCAGAACTTTAGTCGTCGAAGATCGGAGTCGCGACTGGGCTGTTGCGGT 120

Db 61 TGTGCCAGAACTTTAGTCGTCGAAGATCGGAGTCGCGACTGGGCTGTTGCGGT 120

QY 121 GTTGCTGGCCAGCTGTTGGTAAAGTAGTGTGCGAGTTGCAATGGCCATTTGTTGCT 180

Db 121 GTTGCTGGCCAGCTGTTGGTAAAGTAGTGTGCGAGTTGCAATGGCCATTTGTTGCT 180

QY 181 GTTTGGTGGATTAATTCGTTCTAAATCGTCAATCGATGATCAGCAGGATAAATA 240

Db 181 GTTTGGTGGATTAATTCGTTCTAAATCGTCAATCGATGATCAGCAGGATAAATA 240

QY 241 AAGCTAAACCCAGAGTTGGAAGGTAAAGCAGGCAAGTCACACGTTGGCGTAATCCA 300

Db 241 AAGCTAAACCCAGAGTTGGAAGGTAAAGCAGGCAAGTCACACGTTGGCGTAATCCA 300

QY 301 GATACAGCAATAGTTATAGTGTTCAGCCAGTTCGCTACTTACCAGCGTTACAAAGCAA 360

Db 301 GATACAGCAATAGTTATAGTGTTCAGCCAGTTCGCTACTTACCAGCGTTACAAAGCAA 360

QY 361 GAGCGTGGCCAGCATATTGTCGAGAAATTCAGCAAGGCGGATGATTCAGGCGAGAG 420

Db 361 GAGCGTGGCCAGCATATTGTCGAGAAATTCAGCAAGGCGGATGATTCAGGCGAGAG 420

QY 421 CAAGAGATTACGGCAGCTGATCGCGGCAACCGGATGTTGCGTGGCAAGTCAATTCACA 480

Db 421 CAAGAGATTACGGCAGCTGATCGCGGCAACCGGATGTTGCGTGGCAAGTCAATTCACA 480

QY 481 GAAAAA 486

Db 481 GAAAAA 486

RESULT 2

ABK52401

ID ABK52401 standard; DNA; 486 BP.

XX AC

XX ABK52401;

XX 12-AUG-2002 (first entry)

XX DNA encoding outer surface lipoprotein Ospa.

XX Outer surface lipoprotein; Ospa; antibacterial; immunosuppressive;

XX vaccine; poikilothermic fish; fin-fish; Rickettsial septicemia;

XX Rickettsial disease; gene; ds.

XX *Piscirickettsia salmonis*.

XX Key Location/Qualifiers

XX CDS 1..486

XX /tag= a

XX /product= "Ospa"

XX /note= "Outer surface lipoprotein"

XX /partial= "No stop codon given"

XX CA2339327-Al.

XX 15-MAR-2002.

XX 19-MAR-2001; 2001CA-2339327.

XX 15-SEP-2000; 2000US-0677374.

XX (THOR/) THORNTON J C.

XX (KAYW/) KAY W W.

XX (BURL/) BURIAN J.

XX (KUZU/) KUZUK M A.

XX Thornton JC, Kay WW, Burian J, Kuzyk MA;

XX WPI; 2002-455221/49.

XX P-PSDB; AAU97867.

XX Inducing immunity in fin fish to Rickettsial septicemia, comprises
PT administration of an outer surface lipoprotein (Ospa) of a bacterial
PT strain, as a vaccine

XX Claim 19; Fig 2; 55pp; English.

XX The invention describes a method of protecting a poikilothermic fish
CC against infection by the bacterial pathogen *Piscirickettsia salmonis*
CC comprising administering either intraperitoneally, by immersion or
CC orally, an immunogenic amount of principal antigen, the Ospa (outer
CC surface lipoprotein), its variants, non-lipidated form or antigenic
CC peptides derived or synthesized with or without an adjuvant. The new
CC method is used to provide an outer surface lipoprotein (Ospa) of
CC bacterial strain *Piscirickettsia salmonis* as a vaccine to induce immunity
CC in fin-fish against Rickettsial septicemia and other related
CC Rickettsial diseases caused by either a virus, bacteria or parasite.
CC This sequence encodes the *Piscirickettsia salmonis* outer surface
CC lipoprotein, Ospa, used in the creation of the vaccine described in the
CC invention.

XX Sequence 486 BP; 137 A; 79 C; 144 G; 126 T; 0 other;

Query Match 100.0%; Score 486; DB 24; Length 486;
Best Local Similarity 100.0%; Pred. No. 4e-137;
Matches 486; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGAACAGAGGATGTTGCAAGTAGTCTAATATTATTATCAGTGTGTTAGTTGGC 60

Db 1 ATGAACAGAGGATGTTGCAAGTAGTCTAATATTATTATCAGTGTGTTAGTTGGC 60

QY 61 TGTGCCAGAACTTTAGTCGTCGAAGATCGGAGTCGCGACTGGGCGTGTGTTGCGGT 120

Db 61 TGTGCCAGAACTTTAGTCGTCGAAGATCGGAGTCGCGACTGGGCGTGTGTTGCGGT 120

QY 121 GTTGCTGGCCAGCTGTTGGTAAAGTAGTGTGCGAGTTGCAATGGCCATTTGTTGCT 180

```

Db 121 GTTGTGGCCAGCTGTTTGGTAAGGTAGTGGTGGAGTTCAGTGGCCATGGTGGTGGCT 180
QY 181 GTTGTGGTGGATTAATGTTCTTAAATCGGTCAATCGATGATGATGATGATGATGATGATGAT 240
Db 181 GTTGTGGTGGATTAATGTTCTTAAATCGGTCAATCGATGATGATGATGATGATGATGATGAT 240
QY 241 AAGCTAACACAGAGTTTGGAAAGGTAAAGCAGGCGCAAGTGACAGCTTGGCGTAAATCCA 300
Db 241 AAGCTAACACAGAGTTTGGAAAGGTAAAGCAGGCGCAAGTGACAGCTTGGCGTAAATCCA 300
QY 301 GATACAGGCAATAGTTATAGTGTGGCCAGTGCCTACTTACCAGCGTTTACAAATAAGCAA 360
Db 301 GATACAGGCAATAGTTATAGTGTGGCCAGTGCCTACTTACCAGCGTTTACAAATAAGCAA 360
QY 361 GAGCGTCCGAGCAATATTCGAGATTTTCAGCAAAAGCGGATGATGATGATGATGATGATGATGAT 420
Db 361 GAGCGTCCGAGCAATATTCGAGATTTTCAGCAAAAGCGGATGATGATGATGATGATGATGATGAT 420
QY 421 CAAGAGATTACGCACTGATGCGGCAACCGGATGGTGGTGGCAAGTCAATTTCAACA 480
Db 421 CAAGAGATTACGCACTGATGCGGCAACCGGATGGTGGTGGCAAGTCAATTTCAACA 480
QY 481 GAAAAA 486
Db 481 GAAAAA 486

```

RESULT 3

AAH79040
ID AAH79040 standard; cdna; 489 BP.

XX AC AAH79040;

XX 15-JAN-2002 (first entry)

XX Piscirickettsia salmonis polynucleotide P10.6.

XX Piscirickettsia salmonis; piscirickettsiosis; salmonid rickettsial;
 KW septicemia; SRS; surface antigen; vaccine; antibacterial; fish;
 KW ATCC VR-1361; ss.

XX Piscirickettsia salmonis.

XX Key Location/Qualifiers
 FT CDS 1..489
 FT /*tag= a

XX WO200168865-A2.

XX 20-SEP-2001.

XX 12-MAR-2001; 2001WO-GB01055.

XX 11-MAR-2000; 2000GB-0005838.

XX 01-JUL-2000; 2000GB-0016080.

XX 01-JUL-2000; 2000GB-0016082.

XX 29-JUL-2000; 2000GB-0018599.

XX (AQUA-) AQUA HEALTH EURO LTD.

XX Smard N, Brouwers H, Jones S, Griffiths S, Valenzuela P;
 PI Burzio L;

XX WPI; 2001-639050/73.

XX P-PSDB; AAG78025.

XX New nucleic acids encoding an amino acid sequence homologous to the
 PT surface antigen present on piscirickettsia salmonis are useful to
 PT protect fish against piscirickettsiosis -

XX Claim 3; Fig 4; 25pp; English.

XX

CC The invention relates to nucleic acid sequences and the encoded protein
 CC of a least part of the surface antigen present on *Piscirickettsia*
 CC salmonis for production of a vaccine with antibacterial activity to
 CC protect fish against *P. salmonis* which causes piscirickettsiosis, also
 CC known as salmonid rickettsial septicemia.

XX Sequence 489 BP; 139 A; 79 C; 144 G; 127 T; 0 other;

Query Match 100.0%; Score 486; DB 22; Length 489;
 Best Local Similarity 100.0%; Pred. No. 4e-137;
 Matches 486; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGACAGAGGATGTTGCAAGGTAGTAGTCTAATTAATTAATCAAGTGTGTTTAGTTGGC 60

Db 1 ATGAACAGAGGATGTTTGAAGGTAGTAGTCTAATTAATTAATCAAGTGTGTTTAGTTGGC 60

QY 61 TGTGCCCAGAACATTTAGTCGTCGAAGTCGAGCTCGAGTCGAGTCGGGCTGTTGTTGGCGGT 120

Db 61 TGTGCCCAGAACATTTAGTCGTCGAAGTCGAGTCGAGTCGGGCTGTTGTTGGCGGT 120

QY 121 GTTGTGCGCCAGCTGTTTGGTAAAGTGTGTCGAGTTCGAATGGCCATTTGGTGTGCT 180

Db 121 GTTGTGCGCCAGCTGTTTGGTAAAGTGTGTCGAGTTCGAATGGCCATTTGGTGTGCT 180

QY 181 GTTTTGGTGGATTAATTTGTTCTAAATCGGTCAATCGATGATGATGATGATGATGATGAT 240

Db 181 GTTTTGGTGGATTAATTTGTTCTAAATCGGTCAATCGATGATGATGATGATGATGATGAT 240

QY 241 AAGCTAACACAGAGTTTGGAAAGGTAAAGCAGGCGCAAGTGACAGCTTGGCGTAAATCCA 300

Db 241 AAGCTAACACAGAGTTTGGAAAGGTAAAGCAGGCGCAAGTGACAGCTTGGCGTAAATCCA 300

QY 301 GATACAGGCAATAGTTATAGTGTGACCCAGTGCCTACTTACCAGCGTTTACAAATAAGCAA 360

Db 301 GATACAGGCAATAGTTATAGTGTGACCCAGTGCCTACTTACCAGCGTTTACAAATAAGCAA 360

QY 361 GAGCGTCCGAGCAATATTCGAGATTTTCAGCAAAAGCGGATGATGATGATGATGATGATGAT 420

Db 361 GAGCGTCCGAGCAATATTCGAGATTTTCAGCAAAAGCGGATGATGATGATGATGATGATGAT 420

QY 421 CAAGAGATTACGCACTGATGCGGCAACCGGATGGTGGTGGCAAGTCAATTTCAACA 480

Db 421 CAAGAGATTACGCACTGATGCGGCAACCGGATGGTGGTGGCAAGTCAATTTCAACA 480

QY 481 GAAAAA 486

Db 481 GAAAAA 486

RESULT 4

AAF86247
ID AAF86247 standard; DNA; 483 BP.

XX AC AAF86247;

XX 11-JUL-2001 (first entry)

XX DNA sequence of *E. coli* optimised ospA gene 17E2.

XX Poikilothermic fish; piscirickettsia salmonis; rickettsial pathogen;
 KW vaccine; OspA; salmonid rickettsial septicemia; rickettsial disease;
 KW SRS; 17E2; ds.

XX Piscirickettsia salmonis.

OS Synthetic.

XX Key Location/Qualifiers

FT CDS 1..483

FT /*tag= a

FT /partial

FT /product= "OspA"

FT /note= "Genus specific 17kDa antigen, the sequence does
 not include a stop codon"

| | | | |
|----------|---|---|-----|
| Db | 5 | GTGGTTGCCCTCAGGCGACGCTCTCTGATCAATATCTCTGTTTCTCTGGTGGTTGGCGCCC | 64 |
| QY | 68 | AGAACTTTTGTAGTTCGTCGAAGAAGTCGAGAGTCGGGCTGTTGTTGGGGGTTTCTG | 127 |
| Db | 65 | AGAACTTCAGCGCCGAGGAAGTTGCGCGGCCACCGGTGCGGTTGTGGCGGTTGTCGG | 124 |
| QY | 128 | GCCAGCTGTTGGTAAAGTAGTGTGTCAGTTGCAATGGCCATTTGGTGTGCTGTTTTGG | 187 |
| Db | 125 | GCCAGCTGTTGGTAAAGGCTCTGTTGTCGATGGCCATCGGGGTCGGGTTCTGG | 184 |
| QY | 188 | GTGGATTAATTTGGTCTCTAAATCGCTCAATCGATGGATCAGCAGGATAAATAAGCTAA | 247 |
| Db | 185 | GCGGTCGATTGGCTCTAAATCGCTCAGCATGGACGACGAGATAAATAACAATGA | 244 |
| QY | 248 | ACCAGAGTTTGGAAAGGTAAAGCAGGCGCAAGTCACACGTTGGCGTAAATCCAGATACAG | 307 |
| Db | 245 | ACCAGTCTCTGGAAAAGTGAAGCGCCAGGTTACTCGTTGGCGTAATCCGGACACCG | 304 |
| QY | 308 | GCAATAGTTATAGTGTGAGCCAGTGCCTACTTACCGCGTTTCAATAAGCAAGAGCGTC | 367 |
| Db | 305 | GTAAACAGCTACTCTGTGGAACCGGTTCCGACCTACCAAGCTTACAACAACAGGAAGGCC | 364 |
| QY | 368 | GCCAGCAATATCTCGAGAAATTCAGCAAAAGGCCGATGATTCAGGCGCAGAAAGAGA | 427 |
| Db | 365 | GTCAAGCAGTACTGCGCGCAATTCAGCAAAAGCCGATGATTCAGGCGCAGAAAGAGA | 424 |
| QY | 428 | TTTACGGCACTGCATGCGCGCAACCGGATGTTGCGTTGGCAAGTCATTTCAACAGAAAA | 486 |
| Db | 425 | TCTACGGCACCGGTGCGCTCAGCGGATGTCGCGGTGCGCGTGCAGCTGATTAGCACCG | 483 |
| RESULT 6 | | | |
| AAF86248 | | | |
| ID | AAF86248 standard; DNA; 768 BP. | | |
| XX | AAF86248; | | |
| XX | 11-JUL-2001 (first entry) | | |
| DE | DNA sequence of cl7e2 ospA construct with N-terminal fusion partner. | | |
| XX | Poikilothermic fish; Piscirickettsia salmonis; rickettsial pathogen; | | |
| KW | vaccine; OspA; salmonid rickettsial septicaemia; rickettsial disease; | | |
| KW | SRS; l7E2; fusion construct; ds. | | |
| OS | Piscirickettsia salmonis. | | |
| OS | Synthetic. | | |
| XX | Key | | |
| FT | Location/Qualifiers | | |
| FT | 1..768 | | |
| FT | /tag= a | | |
| FT | /partial | | |
| FT | /product= "cl7E2 OspA with N-terminal fusion partner" | | |
| FT | /note= "No stop codon is given" | | |
| FT | misc_feature | | |
| FT | 1..285 | | |
| FT | /tag= b | | |
| FT | /note= "DNA encoding undefined N-terminal fusion partner" | | |
| FT | misc_feature | | |
| FT | 286..768 | | |
| FT | /tag= c | | |
| FT | /note= "Optimised OspA construct cl7E2" | | |
| XX | CA2281913-A1. | | |
| XX | 17-MAR-2001. | | |
| XX | 17-SEP-1999; | | |
| XX | 99CA-2281913. | | |
| XX | 17-SEP-1999; | | |
| XX | 99CA-2281913. | | |
| XX | (KAYW/) KAY W W. | | |
| PA | (BURL/) BURIAN J. | | |
| PA | (KUZV/) KUZV M A. | | |

58

DT 12-AUG-2002 (first entry)

XX DE Escherichia coli codon optimised OspA 17e2, oligonucleotide #5.
 XX KW Outer surface lipoprotein; OspA; antibacterial; immunosuppressive;
 KW vaccine; poikilothermic fish; fin-fish; Rickettsial septicaemia;
 KW Rickettsial disease; 17e2; ss.
 XX OS Synthetic.
 XX PN CA2339327-A1.
 XX PD 15-MAR-2002.
 XX PF 19-MAR-2001; 2001CA-2339327.
 XX PR 15-SEP-2000; 2000US-0677374.
 XX PA (THOR/) THORNTON J C.
 PA (KAYW/) KAY W W.
 PA (BURL/) BURIAN J.
 PA (KUZY/) KUZYK M A.
 XX PI Thornton JC, Kay WW, Burian J, Kuzyk MA;
 XX WP1; 2002-455221/49.
 XX PT Inducing immunity in fin fish to Rickettsial septicaemia, comprises
 PT administration of an outer surface lipoprotein (OspA) of a bacterial
 PT strain, as a vaccine -
 XX PS Example 3; Fig 4; 55pp; English.
 XX CC The invention describes a method of protecting a poikilothermic fish
 CC against infection by the bacterial pathogen *Piscirickettsia salmonis*
 CC comprising administering either intraperitoneally, by immersion or
 CC orally, an immunogenic amount of principal antigen, the OspA (outer
 CC surface lipoprotein), its variants, non-lipidated form or antigenic
 CC peptides derived or synthesized with or without an adjuvant. The new
 CC method is used to provide an outer surface lipoprotein (OspA) of
 CC bacterial strain *Piscirickettsia salmonis* as a vaccine to induce immunity
 CC in fin-fish against Rickettsial septicaemia and other related
 CC Rickettsial diseases caused by either a virus, bacteria or parasite.
 CC This sequence represents an oligonucleotide used to construct an
 CC *Escherichia coli* codon optimised outer surface lipoprotein OspA (17e2)
 CC used in the creation of the vaccine described in the invention.
 XX SQ Sequence 102 BP; 14 A; 23 C; 30 G; 35 T; 0 other;
 Query Match 14.2%; Score 69; DB 24; Length 102;
 Best Local Similarity 80.2%; Pred. No. 4.4e-11;
 Matches 81; Conservative 0; Mismatches 20; Indels 0; Gaps 0;
 QY 337 ACTTACCAGCGTTACAAATAAGCAAGAGCGTCGCCAGCAATATTGTCGAGAAATTCAGCAA 396
 DB 102 ACCTACCAGCGTTACAAACACAGGACGCCGTGACGAGTACTGCCGCGAATTCAGCAG 43
 QY 397 AAGCGGATGATTGCGGCGCAAGCAAGAGATTTCAGGCAC 437
 DB 42 AAAGCCATGATGCGCAGGTCAGAAACAGGAAATCTACGGCAC 2
 RESULT 14
 AAF86253
 ID AAF86253 standard; DNA; 94 BP.
 XX AC AAF86253;
 XX DX 11-JUL-2001 (first entry)
 XX DE PCR primer #3 used in cloning an optimisation of OspA gene.
 XX KW Poikilothermic fish; *Piscirickettsia salmonis*; rickettsial pathogen;
 KW vaccine; OspA; salmonid rickettsial septicaemia; rickettsial disease;

KW SRS; 17E2; fusion construct; PCR primer; ss.
 XX *Piscirickettsia salmonis*.
 XX PN CA2281913-A1.
 XX PD 17-MAR-2001.
 XX PF 17-SEP-1999; 99CA-2281913.
 XX PR 17-SEP-1999; 99CA-2281913.
 XX PA (KAYW/) KAY W W.
 PA (BURL/) BURIAN J.
 PA (KUZY/) KUZYK M A.
 XX PI Kay WW, Burian J, Kuzyk MA;
 XX WP1; 2001-316844/34.
 XX PT Method for protecting poikilothermic fish against salmonid rickettsial
 PT septicaemia and other rickettsial diseases comprises administering a
 PT vaccine containing the OspA protein of *Piscirickettsia salmonis* -
 XX PS Example 3; Fig 4B; 35pp; English.
 XX CC This invention relates to a method for the protection against infection
 CC of a poikilothermic fish by the bacterial pathogen, *Piscirickettsia*
 CC *salmonis*. The method comprises administering an immunogenic amount of a
 CC *P. salmonis* specific antigen termed OspA, or an immunogenic fragment of
 CC OspA in the form of a vaccine. The method is used for protecting animals,
 CC particularly poikilothermic fish, against the bacterial pathogen
 CC *P. salmonis*. The method is also useful for protecting against salmonid
 CC rickettsial septicaemia (SRS) and other rickettsial diseases. The present
 CC sequence represents a PCR primer used in the cloning and optimisation of
 CC the *P. salmonis* OspA gene. The OspA gene is used in the method of the
 CC invention.
 XX SQ Sequence 94 BP; 18 A; 20 C; 35 G; 21 T; 0 other;
 Query Match 12.8%; Score 62; DB 22; Length 94;
 Best Local Similarity 78.7%; Pred. No. 5.6e-09;
 Matches 74; Conservative 0; Mismatches 20; Indels 0; Gaps 0;
 QY 142 AAAGGTAGTGTGCGATGTCGAATGCCATGGTGGTGTGCTGTTTGGTGGATTAAATGGT 201
 DB 1 AAAGGCTCTGCTGCTGTGGCGATGCCATCGCGGTCTCTGGCGGCTCTGATTGGC 60
 QY 202 TCTAAATCGGTCAATCGATCGATCAGCAGGATA 235
 DB 61 TCTAAATCGGTCAAGATCGATCGACGAGGATA 94
 RESULT 15
 ABK52408
 ID ABK52408 standard; DNA; 94 BP.
 XX AC ABK52408;
 XX DX 12-AUG-2002 (first entry)
 XX DE *Escherichia coli* codon optimised OspA 17e2, oligonucleotide #3.
 XX KW Outer surface lipoprotein; OspA; antibacterial; immunosuppressive;
 KW vaccine; poikilothermic fish; fin-fish; Rickettsial septicaemia;
 KW Rickettsial disease; 17e2; ss.
 XX OS Synthetic.
 XX PN CA2339327-A1.
 XX PD 15-MAR-2002.

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 22, 2003, 13:11:08 ; Search time 1355.32 Seconds
(without alignments)
10435.915 Million cell updates/sec

Title: US-09-677-374-1

Perfect score: 486

Sequence: 1 atgaacagagatgttgca.....aagtcatttcacagaaaaa 486

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.*

1: gb_ba.*

2: gb_hgt.*

3: gb_in.*

4: gb_om.*

5: gb_ov.*

6: gb_pat.*

7: gb_ph.*

8: gb_pl.*

9: gb_pr.*

10: gb_ro.*

11: gb_sts.*

12: gb_sy.*

13: gb_un.*

14: gb_vt.*

15: em_ba.*

16: em_fun.*

17: em_hum.*

18: em_in.*

19: em_mu.*

20: em_om.*

21: em_or.*

22: em_pat.*

23: em_ph.*

24: em_pl.*

25: em_ro.*

26: em_sts.*

27: em_un.*

28: em_vt.*

29: em_vt.*

30: em_hgt_hum.*

31: em_hgt_inv.*

32: em_hgt_other.*

33: em_hgt_mus.*

34: em_hgt_pln.*

35: em_hgt_rod.*

36: em_hgt_mam.*

37: em_hgt_vrt.*

38: em_sy.*

39: em_hgtgo_hum.*

40: em_hgtgo_mus.*

41: em_hgtgo_other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Match % | Length | DB | ID | Description |
|------------|-------|---------|--------|----|------------|-----------------|
| 1 | 486 | 100.0 | 489 | 6 | AX252413 | Sequence |
| 2 | 486 | 100.0 | 4983 | 1 | AF184152 | Piscirickettsia |
| 3 | 79 | 16.3 | 448 | 1 | RPX004 | Rickettsia |
| 4 | 70.6 | 14.5 | 237523 | 1 | RPX004 | Rickettsia |
| 5 | 68.4 | 14.1 | 416 | 1 | AF031534 | Rickettsia |
| 6 | 65.8 | 13.5 | 535 | 1 | RIRANT17KB | Rickettsia |
| 7 | 65.8 | 13.5 | 537 | 1 | RIR17KCA | Rickettsia |
| 8 | 64.6 | 13.3 | 539 | 1 | RIRANT17KA | Rickettsia |
| 9 | 64.6 | 13.3 | 10127 | 1 | AE008675 | Rickettsia |
| 10 | 64.2 | 13.2 | 539 | 1 | RIRANT17KD | Rickettsia |
| 11 | 64.2 | 13.2 | 620 | 1 | RIRANTRR | Rickettsia |
| 12 | 62.2 | 12.8 | 434 | 1 | MRI269516 | Male-kill |
| 13 | 62.2 | 12.8 | 434 | 1 | MRI269517 | Male-kill |
| 14 | 62.2 | 12.8 | 434 | 1 | MRI269518 | Male-kill |
| 15 | 59.4 | 12.2 | 546 | 1 | AF195118 | Rickettsia |
| 16 | 59 | 12.1 | 489 | 1 | RRU11020 | Rickettsia |
| 17 | 58.2 | 12.0 | 415 | 1 | AF027124 | Rickettsia |
| 18 | 57.8 | 11.9 | 532 | 1 | RIRANT17KC | Rickettsia |
| 19 | 57.6 | 11.9 | 411 | 1 | AF060704 | Rickettsia |
| 20 | 57.6 | 11.9 | 411 | 1 | AF060706 | Rickettsia |
| 21 | 57.4 | 11.8 | 489 | 1 | RP017008 | Rickettsia |
| 22 | 57.4 | 11.8 | 498 | 1 | RAU11013 | Rickettsia |
| 23 | 56.2 | 11.6 | 454 | 1 | AF181036 | Rickettsia |
| 24 | 55.6 | 11.4 | 489 | 1 | RMU11017 | Rickettsia |
| 25 | 54.4 | 11.2 | 395 | 1 | U04162 | Rickettsia |
| 26 | 54.2 | 11.2 | 492 | 1 | RIRTRAPRO | Rickettsia |
| 27 | 52.6 | 10.8 | 491 | 1 | RIRGENSEQ | Rickettsia |
| 28 | 49 | 10.1 | 23093 | 2 | AC114815 | Homo sapi |
| 29 | 47.8 | 9.8 | 394 | 1 | AF260571 | Rickettsia |
| 30 | 47.8 | 9.8 | 7218 | 6 | 166494 | Sequence |
| 31 | 46.6 | 9.6 | 206114 | 2 | AC118508 | Rattus no |
| 32 | 46.4 | 9.5 | 111824 | 9 | AL160412 | Human DNA |
| 33 | 46.4 | 9.5 | 206114 | 2 | AC118508 | Rattus no |
| 34 | 46.2 | 9.5 | 165087 | 2 | AC102004 | Mus muscu |
| 35 | 45.4 | 9.3 | 68651 | 9 | AC027343 | Homo sapi |
| 36 | 45.4 | 9.3 | 121829 | 2 | AC103214 | Rattus no |
| 37 | 45.4 | 9.3 | 152198 | 2 | AC128636 | Rattus no |
| 38 | 45.2 | 9.3 | 162165 | 2 | AC127626 | Rattus no |
| 39 | 45 | 9.3 | 61052 | 2 | AC117074 | Dictyoste |
| 40 | 45 | 9.3 | 61052 | 2 | AC123513 | Dictyoste |
| 41 | 45 | 9.3 | 156370 | 2 | AL158162 | Homo sapi |
| 42 | 45 | 9.3 | 166464 | 9 | AL158212 | Human DNA |
| 43 | 44.8 | 9.2 | 212383 | 2 | AC119030 | Rattus no |
| 44 | 44.8 | 9.2 | 217416 | 2 | AC103538 | Rattus no |
| 45 | 44.6 | 9.2 | 13123 | 1 | AE003872 | Xylella f |

ALIGNMENTS

RESULT 1

AX252413

LOCUS

DEFINITION

AX252413

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

AX252413 Sequence 5 from Patent WO0168865. 489 bp DNA linear PAT 05-OCT-2001

AX252413.1 GI:15985721

Piscirickettsia salmonis.

Piscirickettsia salmonis

Bacteria; Proteobacteria; gamma subdivision; Piscirickettsia group;

Piscirickettsia

1 (bases 1 to 489)

Simard,N., Brouwers,H., Jones,S.F., Griffiths,S., Valenzuela,P. and

Burzio,L.

Fish vaccine against piscirickettsia salmonis

GNSYSVEPVRTYQRYNKQERROQYCREFOOKAMIAQOEIYGTACRQPDGRQWVIST
EX*
repeat_region 3560..3846
gene complement(3842..4561)
/gene="tnpA"
CDS complement(3842..4561)
/gene="tnpA"
/codon_start=1
/transl_table=11
/product="transposase"
/protein_id="AAG16999.1"
/db_xref="GI:10411345"
/translation="MDLFLISLFCVIDDCEQLLPQWAILLEDINKKRNKPSQMS
EIMTMIYFHKSNYRNFMYLHVINGSWVKYFPNSVSNREVELMPSILLPLCFE
AQGTATGTYFYDSILRVCHEKRAKSONRAKSLKSTMGWYGFKLHIVNDMG
ELMAFKMSKATDDRVLPKMAENLTGLIGDKYISQKLFQLYEKGILQITKIRKN
MENKLVIKIDKILLKRGIIA"
repeat_region 4562..4849
BASE COUNT 1362 a 778 c 1265 g 1578 t
ORIGIN
Query Match 100.0%; Score 486; DB 1; Length 4983;
Best Local Similarity 100.0%; Pred. No. 1.8e-129;
Matches 486; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ATGAACAGAGGATGTTGCAAGTAGTAGTCTAATATTATTCAGTGTGTTTTAGTTGGC 60
|||||
Db 2834 ATGAACAGAGGATGTTGCAAGTAGTAGTCTAATATTATTCAGTGTGTTTTAGTTGGC 2893
QY 61 TGTGCCAGAACTTTAGTCTCAAGAAGTCGGAGTCGCGACTGGGCTGTGTTGGCGGT 120
|||||
Db 2894 TGTGCCAGAACTTTAGTCTCAAGAAGTCGGAGTCGCGACTGGGCTGTGTTGGCGGT 2953
QY 121 GTTGCTGGCCAGCTGTTGGTAAAGTAGTGTGCGAGTTGCAATGGCCATTGGTGTGCT 180
|||||
Db 2954 GTTGCTGGCCAGCTGTTGGTAAAGTAGTGTGCGAGTTGCAATGGCCATTGGTGTGCT 3013
QY 181 GTTTGGTGGGTAATTTGTTCTAAATCGGTCAATCGATGGATCAGCAGGATAAATA 240
|||||
Db 3014 GTTTGGTGGGTAATTTGTTCTAAATCGGTCAATCGATGGATCAGCAGGATAAATA 3073
QY 241 AAGCTAAACAGAGTTTGGAAAAGTTAAAGCAGGCAAGTGACACGTTGGCGGTAAATCCA 300
|||||
Db 3074 AAGCTAAACAGAGTTTGGAAAAGTTAAAGCAGGCAAGTGACACGTTGGCGGTAAATCCA 3133
QY 301 GATACAGGCAATAGTTATAGTTGTAGCCAGTGCCTACTTTACCAGCGTTACAATAAGCAA 360
|||||
Db 3134 GATACAGGCAATAGTTATAGTTGTAGCCAGTGCCTACTTTACCAGCGTTACAATAAGCAA 3193
QY 361 GAGCGTGGCCAGCAATATTTGCGAGATTTTCAGCAAAAGCGGATGATTGAGGCGCAGAAG 420
|||||
Db 3194 GAGCGTGGCCAGCAATATTTGCGAGATTTTCAGCAAAAGCGGATGATTGAGGCGCAGAAG 420
QY 421 CAAGAGATTTACGCGACTGATCGCGCAACCGGATGCTGTTGGCAAGTCATTTCACCA 480
|||||
Db 3254 CAAGAGATTTACGCGACTGATCGCGCAACCGGATGCTGTTGGCAAGTCATTTCACCA 3313
QY 481 GAAAAA 486
|||||
Db 3314 GAAAAA 3319
RESULT 3
RSU76907
LOCUS RSU76907 448 bp DNA linear BCT 18-FEB-1998
DEFINITION Rickettsia sp. 17kDa common-antigen gene, partial cds.
ACCESSION U76907
VERSION U76907.1 GI:2894794
KEYWORDS Rickettsia sp.
SOURCE Rickettsia sp.
ORGANISM Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
Rickettsiaceae; Rickettsieae; Rickettsia.

REFERENCE 1 (bases 1 to 448)
Davis, M.J., Ying, Z., Brunner, B.R., Pantoja, A. and Ferwerda, F.H.
Rickettsial relative associated with papaya bunchy top disease
Curr. Microbiol. 36 (2), 80-84 (1998)
JOURNAL MEDLINE 98087556
PUBMED 9425244
REFERENCE 2 (bases 1 to 448)
Ying, Z. and Davis, M.J.
Direct Submission
Submitted (01-NOV-1996) Tropical Research and Education Center,
University of Florida, 18905 SW 280 Street, Homestead, FL 33011,
USA
FEATURES
source Location/Qualifiers
1..448
/organism="Rickettsia sp."
/db_xref="taxon:789"
/note="Puerto Rico"
CDS <1..>448
/note="putative"
/codon_start=3
/transl_table=11
/product="17 kDa common-antigen"
/protein_id="AAC02809.1"
/db_xref="GI:2894795"
/translation="ALAASMLEACQPGGKMGKGGSTLIGTLGLVGSQFGGTGRL
AAGVAGALLGAILGNOIGAGMDEQDRKLAELTSQRALEAAPSGSSVQWRNPONGNYGT
VTPSKAYKNNTGQYCYREYQTIVVGGKQOKAYGTACRQPDGQMVV"
BASE COUNT 135 a 81 c 120 g 112 t
ORIGIN
Query Match 16.3%; Score 79; DB 1; Length 448;
Best Local Similarity 51.8%; Pred. No. 8.5e-12;
Matches 206; Conservative 0; Mismatches 190; Indels 3; Gaps 1;
QY 72 CTTTAGTCGCAAGAAGTCGGAGTCGCGACTGGGCTGTGTTGGCGGTCTGTGTCGCCA 131
|||||
Db 47 CATGAATAACAAGTAGTAGTACGCTAATTTGGCGGAACTTTAGGAGCACTTTGTGTTTC 106
QY 132 GCTGTTTGGTAAAGTAGTGTGCGAGTTGCAATGGCCATTGGTGTGCTGTTTGGGTGG 191
|||||
Db 107 ACAATTTGGTGGTACGGGACGACTTGTGCGGTAGGAGCAGTGCCTTACTTGTGTCG 166
QY 192 ATTAATTTGGTCTTAAATTCGGTCAATCGATGGATCAGCAGGNTAAAATAAGCTAAACCA 251
|||||
Db 167 AATCCTTGGCAACCAATTTGTCAGGTATGGATGAACAAGATAGAAAACCTTGCAGAGCT 226
QY 252 GAGTTTGGAAAAGTAAAGCAGGCAAGTGACACGTTGGCGTAAATCCAGATACAGSCAA 311
|||||
Db 227 TACTTCTCAAGAGCTTTAGAAG---CTGCACCAAGCGGTAGCAGTGTACAATGGCGTAA 283
QY 312 TAGTTATAGTGTGAGCCAGTGTGCTACTTACCAGCGTTACAATAAGCAAGAGCGTCGCCA 371
|||||
Db 284 TCCTGATAATGGTAAATTTATGTTACTGTGACACCAAGCAAAAGCTTATAAAAATAATACCGG 343
QY 372 GCATATTTGCGAGATTTTCAGCAAAAGCGGATGTTGAGGCGCAGNAGCAAGATTTA 431
|||||
Db 344 TCAATATTTGCGTGAATACACCAACAGTTGTAGTAGTGGAAAAACAACAAAAAGCTTA 403
QY 432 CGGCACCTGTCATGCCGCAACCGGATGCTGTTGGCAAGT 470
|||||
Db 404 TGGCACTGCTCGCTCAACCTGATGGACAATGGCAAGT 442
RESULT 4
RPXX04
LOCUS RPXX04 237523 bp DNA linear BCT 11-NOV-1998
DEFINITION Rickettsia prowazekii strain Madrid E, complete genome; segment
4/4.
ACCESSION AJ235273
VERSION AJ235273.1 GI:3861237
KEYWORDS complete genome.
SOURCE Rickettsia prowazekii.
ORGANISM Rickettsia prowazekii.

Bacteria; Proteobacteria; alpha subdivision; Rickettsiales; Rickettsiaceae; Rickettsiae; Rickettsia; typhus group.

1 (bases 1 to 237523)
 Andersson, S.G., Zomorodipour, A., Andersson, J.O., Sickeritz-Ponten, T., Alsmark, U.C., Podowski, R.M., Naslund, A.K., Eriksson, A.S., Winkler, H.H. and Kurland, C.G.
 The genome sequence of Rickettsia prowazekii and the origin of mitochondria
 Nature 396 (6707), 133-140 (1998)

99039499
 9823893
 2 (bases 1 to 237523)
 Andersson, S.G.E.
 Direct Submission
 Submitted (11-NOV-1998) S.G.E. Andersson,
 Div. Andersson@molbio.un.se, Dept. of Molecular Biology, University
 of Uppsala, Husargatan 3, Uppsala, S-751 24, SWEDEN
 Location/Qualifiers
 1. .237523
 /organism="Rickettsia prowazekii"
 /strain="Madrid E"
 /db_xref="taxon:782"
 164. .1123
 /gene="Rp701"
 164. .1123
 /gene="Rp701"
 /codon_start=1
 /transl_table=11
 /product="unknown"
 /protein_id="CAA15137.1"
 /db_xref="GI:3861238"
 translation="MAEPIPTLFGNKRDLRINNOKALNKLHIIPLOVNRKSHIS
 AKRIFSELKSKKEKIPFNFQIISNNVYKPIFRAQDLINAKLITITFDKYN
 ALPNSGKSWKYOIEKDVVFTEADQNLSTADGIVPKDKTITIDRLVSVN
 NLIYNHETHELSTIKYKVDLKLKTAKNDSKRIETWPLSVDEANLITAKFGITS
 EDQYGLKLEHIEHLKPNNAENFKYVYSQISQPKDKDKQKLDNSNENFQKVI
 NDSNQLQVEHTSYQDKKQKQSGFLTRIFNFKNVFKQ"
 2074. .3348
 /gene="Rp702"
 2074. .3348
 /gene="Rp702"
 /codon_start=1
 /transl_table=11
 /product="unknown"
 /protein_id="CAA15138.1"
 /db_xref="GI:3861239"
 translation="MTTILVIVIMISLSALFAATETAITASSPGKTHKLKIAGNKR
 AKTVLEKKEKVIKGLLGNLSINTICTTATLFTSLGNGNTIVASGVMAFII
 VFAVVPRAIVAPEQLTKMSTIVIFLKLFPINIALDYIPKIECFIRINLQ
 ISGTEEVGVIEHYHOGGVYKSKRMMLGGILDIRMTVSEIMTHRSIITAINIDL
 ELIITKLSGARTIPLMQDNRNIIIGILNKLKALYENNDKKVDIINTLTPPW
 FIPDNALVQDLHAFRRERNHFACVDEYGLLGLITLEDVIEIEVGPITDEHRLNN
 ETIQESNFTIKGTTTRDINRELDNLSDEDANTAGLIHIIKRIARPNGGEVIKIF
 NFKIILAKIANKDSVKITVLPWTEETISSE"
 3341. .5353
 /gene="Rp703"
 3341. .5353
 /gene="Rp703"
 /codon_start=1
 /transl_table=11
 /product="CYTOCHROME C-TYPE BIOGENESIS PROTEIN CCMF
 (ccmf)"
 /protein_id="CAA15139.1"
 /db_xref="GI:3861240"
 translation="MSKIGNFLLLTTCMLSGILTLLIPLWLDYLLIKKRRNNICLKFF
 PWICTSOVMQATIFYASSLCSIPAFCTLVAFITSDFSQNVFLHSTLKLPIYKI
 SCSSWSSMLLWFLQIVSCCVIFCLDKMLKFSIIILSAIQLLFSGFIVFTSN
 PHMEFVPKSELGNMQLDIALSHPLLYLGVSVAVPTIVCAFLIAVLSRLD
 NMLPGYFGKITQSDRITSRLKIFAGGILEFTTIGISLGSWAYRELNWGLWE
 FDPVENISLFWLSGLMLHHSIIIVTKRNMOWHTIILSIITFLVIFSTLVRSGFI
 NSHAFAPSSERIEYLLVILGIGGLYVIRLNKCSIEKTIYNTFNWNTSCNMT
 IREKAILIGNIFPLSLILICATLYPTIYALFNDKPIIINEEFPKIFIVIPILC
 TIGLFTNSSIKKHIIMLMLSLIITYFILLNKFVSISILATIMTSIFLMLNHHVLLV

KTWYFRNSLKASTSMILGHFGVALLAFSAITMNSLLOSEMDFICEVGTSKTFNAKVT
 LQNIKFAQCKNYRQIAEFPWMSDHNKNNITILKPNRFYIIEQKLSQESDHSYLLYDL
 YAVLSNIDGKRIIHAQYKPMFSFIWIGVILLTAGFVIALIRKNSY"
 complement(7264. .12195)
 /gene="Rp704"
 complement(7264. .12195)
 /gene="Rp704"
 /codon_start=1
 /transl_table=11
 /product="CELL SURFACE ANTIGEN (sca5)"
 /protein_id="CAA15140.1"
 /db_xref="GI:3861241"
 translation="MAQKPNFLKKIISAGLVASTATIVACFSGVAMGAAMQVNRITN
 AAATFDGIGDQAAGANIPVANSVITANANNPTNTPNGHLSLFLDTANDLAVT
 INEDTTIGFTINAAQAKFNFTVAGKILNITQGITVQESNTINAGNALTKVHGG
 AALNADSLGSLITFAAASPGLMFNLTPOAPLTGANSKIVNGNGTILNITG
 FIOVSNDTEAGIKTINTDDCGLFNFSTPDAANTLNQVGGNTINFGIDGTGLVIV
 SKNGATEFNVTCTLGNLKGILIELNTAAVAKLISOGGAANAVIGDNCAGRAAGFI
 VSDNGNAALISQVYAKNAVISOANAGGVTFEHIYDVGLGCTNFKTADSKVILTE
 NSNFGSTNGLNLDQIVVPDKILKNFIDGVKNNGNTAGVITFNANGALVASTDPN
 IAVTININAEAGAGVVELSGIIHAEELRNGGSIKFLADGTIVNGPVNQNALMNNNA
 LAAGSIQDGSATITDIGNGVNAALQHIITLANDSKILALDGANITIGANGGAIHF
 QANGGTIKLTNTONNIVNFDLDTITDKTGVVDASSLTNQTLLTNGSIGTVVANTKT
 LAQNLGSSKILNAGDVAINELVIENNGSVOLNHNLTILTKTINAAOQOIIVAADP
 LNTNTLADGTNLGSAENPLSTHFATKAANADSI LANVGKVNLYANNITNDANVGS
 LHPFSGTSIVSGTVGQGGHKLNNLIDNGTIVKFLDGTTFNGCTKIEKSLIQISN
 NYTTHVESADNTGTLEFVNTDPIITVLNKGAYFVKLQVITSGPNIIVNEIGNVG
 IVHGAANSISFENASLTSLFSPGTPLDVLTIKSTVGNTVDNFNAPIVVSGIDS
 MINNGQIIGDKNIIALSLGSDNSITVNANTLYSGIRTKNNQCTVTLSGMPNPGT
 IYGLGLENGSPKLVQVTTDYNNIGSIIANNVTINDYVTLTTGCTIAGTDFDAKITLG
 SVGNANVRVDSFSDPRSMIVATQANKGTIVYLGALVSNIGSLDTPVASVRFCTGN
 DSGAGLQGNISQNDIFGTNLTILNSNVILGGTTAI NEIDLLTNLILFANGSTSW
 GDNISTITLVNMSGNGVVIADQVNAATTTTITKIODNANANFSGTQAALIQ
 GGARFTLGAPEFAVGTGSMIFVKYELIRDSNQDYVLTNRDNLVNVTVTAVGNSAIAN
 APGVSONISCRLESTAAVNNMLLAKSDVATFVGAIAITDTSAATVTVNLNDTQT
 QDLISNRLGTLRLYLSNAETSVDAGSATGSSVDEAVSYGVWAKPFYNABODKKG
 IAGYKAKTGVVGLDPLASDNLMGAAIGITDKIDKODYKKGDKTDINGLSPLVG
 SOQLVKNFFAQAQNALFTLNKVKSKSQRYFESNGKSKQAKQAGNYDNMTPGGNLFCY
 DYNAMPNVLPMPAGLSYLKSNENYETGTTVANKRINKSFSDRDLIVGAKVAGST
 VNITDIVIPEIHSFVYHVKVNGKLSNSQSLMDGQTAPFISQPDRTAKTSYNIUSANI
 KSDAKMEYGIQYDFNSKTYTAHQGTILKRVNF"
 complement(12953. .13216)
 /gene="Rp705"
 complement(12953. .13216)
 /note="Pseudogene that lacks ATG or GTG initiation codon."
 Homology to guanosine-3',5'-bis(diphosphate)
 3-pyrophosphohydrolase."
 complement(13563. .14507)
 /gene="Rp706"
 complement(13563. .14507)
 /gene="Rp706"
 /codon_start=1
 /transl_table=11
 /product="unknown"
 /protein_id="CAA15141.1"
 /db_xref="GI:3861242"
 translation="MKIRQATKPVIIIGSGPELTDARELFEHNPLGIILFRNRIRK
 NERGEDKAAALIKLIADIKEVLGDNITIAIDQEGWVKRLIAPTIFYDAPAAQITELQ
 TKYNYSTIAKELREVGINLDFAPVADLIHDGADKIIISDRSFKEPEYVPLFLSAIA
 GLOEKVYACIKHPIPGHGRATVDSHIELPIDTSLKLEDDTDFKVKELAYIKLA
 MTAHVITALDPDNPALQIIVIDYIKSNIGFKGLIISDTIEMQALSGSMADITKAL
 DAGDIVLECTGEFNKSEVLGSVAENSIDKFSDLFIS"
 complement(14552. .15025)
 /gene="Rp707"
 complement(14552. .15025)
 /gene="Rp707"
 /codon_start=1
 /transl_table=11
 /product="unknown"
 /protein_id="CAA15142.1"
 /db_xref="GI:3861243"
 translation="MINNKLTKKYVASEVIRIINLIALHKLRYLETPIGLSHYKINN


```

QY 411 AGGCAGAGCAGACAGATTACGGCAGCTGATCGCGGCAACCGGATGTCGTTG 464
    || | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 361 CGGAAACAAACAAAGCATACGGTAAATGATCGCGGCAACCGGATGTCGTTG 414

RESULT 6
RIRANT17KB RIRANT17KB 535 bp DNA linear BCT 26-APR-1993
LOCUS R.typhi 17K genus-common antigen gene, complete cds.
DEFINITION M28481
ACCESSION M28481
VERSION M28481.1 GI:152459
KEYWORDS antigen.
SOURCE R.typhi DNA.
ORGANISM Rickettsia typhi
REFERENCE 1 (bases 1 to 535)
Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
Rickettsiaceae; Rickettsiae; Rickettsia; typhus group.
AUTHORS Anderson, B.E. and Tzianabos, T.
TITLE Comparative sequence analysis of a genus-common rickettsial antigen
gene
JOURNAL J. Bacteriol. 171 (9), 5199-5201 (1989)
MEDLINE 89359171
PUBMED 2768201
COMMENT Draft entry and printed copy of sequence for [1] kindly submitted
by B.Anderson, 27-OCT-1989.
FEATURES
    source
        1..535
            /organism="Rickettsia typhi"
            /db_xref="taxon:785"
        1..6
            /note="17 kD antigen precursor"
        21..27
            /codon_start=1
        37..535
            /product="17 kD mRNA"
        56..535
            /transl_table=11
        116..532
            /protein_id="AAA26377.1"
            /db_xref="GI:152460"
            /translation="MKLLSKVMILALAAASMLQACNPGCMNKQGTGTLGGAGGALLG
            SOFGKGQLVGVGALLGAVLGOIGASLDEODRKLLETSORALESPSGSNIEW
            RNPDNHGYVTPNKTYSNSTGYQCREYTVIIGGKQITGYNACRQPDQWQVNV"
        56..115
            /note="17 kD antigen signal peptide"
        116..532
            /product="17 kD antigen"
    sig_peptide 182 a 93 c 122 g 138 t
    mat_peptide 182 a 93 c 122 g 138 t
BASE COUNT 182 a 93 c 122 g 138 t
ORIGIN
    Query Match 13.5%; Score 65.8; DB 1; Length 535;
    Best Local Similarity 49.9%; Pred. No. 5.8e-08;
    Matches 194; Conservative 0; Mismatches 192; Indels 3; Gaps 1;

QY 82 CAAGAAGTCGGAGCTCGGACTGGGCTGTGTGGCGGTGTGCTGGCCAGCTGTTGGT 141
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 140 CAAGGACTGGGAACACTTCTTGGCGCGCGGCTGTCATTACTTGGTCTCAATTCGGT 199
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 142 AAAGTAGTGGTTCGAATGCGCAATGCGTGTGTGCTGTTTGGGTGGATTAAATGGT 201
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 200 CACGGTAAGGACACTTGTCCGAGTAGGTGTAGCGGCATTACTTGGGCGACTTCTTGGT 259
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 202 TCTAAATCGGTCATPCGATGATCAGCAGGATAAATAAGCTAAACACAGATTGGAA 261
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 260 GGACAAATCGGTGCAAGTCTGGATGAGCAGGATAGAAAACCTTAGAACATACACAA 319
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 262 AAGGTAAGCAGGCGCAAGTGCACGTGGCGTATCCAGATACAGGCAATAGTTATAGT 321
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 320 AGAGCTTTAGAACTCT--GCTCCTAGCGGTAGTACATAGAAATGCGCAATCCAGATAAT 376
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 322 GTTGAGCCAGTCGCTACTTACCAGCGTTTACAATAAGCAAGAGCGTCGCCAGCAATATGCT 381
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 377 GCATCATGGTTAGTACACCTTAATAAACTTTATAGAACAGTACAGGTCAATATGTC. 436

```

```

QY 382 CGAGAAATTCACGAAAGCGGATGATTCAGGCGCAGAACGACAGATTTACGCGACTGCA 441
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 437 CCGTAAATACACTCAACACAGTCTGTAATAGGGGGAACAAACAAACATATGAAATGCA 496
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 442 TGCCGCGCAACCGGATGCTGTTGGCAAGT 470
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 497 TGCCGCGCAACCTGACGGACAATGGCAAGT 525
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 7
RIR17KCCA RIR17KCCA 537 bp DNA linear BCT 04-FEB-1999
LOCUS Rickettsia japonica 17K genus-common antigen gene, complete cds.
DEFINITION D16515
ACCESSION D16515
VERSION D16515.1 GI:695412
KEYWORDS 17K genus-common antigen; 17k dalton protein.
SOURCE Rickettsia japonica (sub_species:YH) DNA.
ORGANISM Rickettsia japonica
REFERENCE 1 (bases 1 to 537)
Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
Rickettsiaceae; Rickettsiae; Rickettsia; spotted fever group.
AUTHORS Furuya, Y., Katayama, T., Yoshida, Y. and Kaiho, I.
TITLE Specific amplification of Rickettsia japonica DNA from clinical
specimens by PCR
JOURNAL J. Clin. Microbiol. 33 (2), 487-489 (1995)
MEDLINE 95229950
PUBMED 75229950
COMMENT 2 (bases 1 to 537)
Yoshida, Y.
AUTHORS Direct Submission
TITLE Submitted (05-JUL-1993) Yoshiya Yoshida, Kanagawa Prefectural
Public Health Laboratory; 52-2 Nakao-cho, Asahi-ku, Yokohama,
Kanagawa 241, Japan (Tel:045-363-1030, Fax:045-363-1037)
FEATURES
    source
        1..537
            /organism="Rickettsia japonica"
            /sub_species="YH"
            /db_xref="taxon:35790"
            /note="propagated in Vero cells"
        58..537
            /note="unnamed protein product"
            /codon_start=1
            /transl_table=11
            /protein_id="BAA03965.1"
            /db_xref="GI:1435056"
            /translation="MKLLSKVIMILALAAASMLQACNPGCMNKQGTGTLGGAGGALLG
            SOFGKGQLVGVGALLGAVLGOIGAGMDEODRRLAELTSORALETSAPSGSNIEW
            RNPDNHGYVTPNKTYSNSTGYQCREYTVIIGGKQKAYGNACRQPDQWQVNV"
        176 a 92 c 129 g 140 t
BASE COUNT 176 a 92 c 129 g 140 t
ORIGIN
    Query Match 13.5%; Score 65.8; DB 1; Length 537;
    Best Local Similarity 49.9%; Pred. No. 5.8e-08;
    Matches 194; Conservative 0; Mismatches 192; Indels 3; Gaps 1;

QY 82 CAAGAAGTCGGAGCTCGGACTGGGCTGTGTGGCGGTGTGCTGGCCAGCTGTTGGT 141
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 142 CAAGGTACAGGAACACTTCTTGGCGGTGTGCTGGCGCATTAATTGTTCTCAATTCGGT 201
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 142 AAAGTAGTGGTTCGAATGCGCAATGCGGCTGTGTGCTGTTTGGGTGGATTAAATGGT 201
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 202 AAGGCAACAGGACAGCTTGTGGAGTAGGTGTAGTGTGCTTACTTGGACAGCTTCTTGGT 261
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 202 TCTAAATCGGTCATPCGATGATCAGCAGGATAAATAAGCTAAACACAGATTGGAA 261
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 262 GGACAAATCGGTGCAAGTATGATGAGCAGGATAGAGACTTGCAGAGCTTACCTCACAG 321
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 262 AAGGTAAGCAGGCGCAAGTGCACGTGGCGTATCCAGATACAGGCAATAGTTATAGT 321
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 322 AGAGCTTTAGAAAC---AGCTCCTAGTGTAGTAACTAGGCTAGGCTAACCGGATAAC 378
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 322 GTTGAGCCAGTCGCTACTTACCAGCGTTTACAATAAGCAAGAGCGTCGCCAGCAATATGCT 381
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |

```



```
Db 379 GCGAATTACGGTTACGTAACACCTAATAAACTTATAGAAATAGCACTGGTCAATATGCG 438
QY 382 CGAGAAATTCAGCAAAAGCGGATGATGTCAGGCGCAGAACGAGATTTACGGCAGTCGCA 441
Db 439 CGTGAGTACACTCAACACAGTTGTAATAGGCGGAAACAAACAAAGACATACGGTAATGCG 498
QY 442 TCGCGGCAACCGGATGTCGTTGGCAAGT 470
Db 499 TCGCGGCAACCTGACGGACAATGGCAAGT 527

RESULT 8
LOCUS RIRANT17KA 539 bp DNA linear BCT 26-APR-1993
DEFINITION R.conorii 17K genus-common antigen gene, complete cds.
ACCESSION M28480
VERSION M28480.1 GI:152457
KEYWORDS antigen.
SOURCE R.conorii DNA.
ORGANISM Rickettsia conorii
Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
Rickettsiaceae; Rickettsieae; Rickettsia; spotted fever group.
REFERENCE 1 (bases 1 to 539)
AUTHORS Anderson, B.E. and Tzianabos, T.
TITLE Comparative sequence analysis of a genus-common rickettsial antigen
gene
JOURNAL J. Bacteriol. 171 (9), 5199-5201 (1989)
MEDLINE 89359171
PUBMED 2768201
COMMENT Draft entry and printed copy of sequence for [1] kindly submitted
by B.Anderson, 27-OCT-1989.
FEATURES
source Location/Qualifiers
1..539 /organism="Rickettsia conorii"
/db_xref="taxon:781"
1..6
-35_signal 21..27
-10_signal 38..>539
mRNA /product="17 kD mRNA"
60..539
CDS /note="17 kD antigen precursor"
/codon_start=1
/transl_table=11
/db_xref="GI:152458"
/translation="MKLLSKIMIIALATSMLOACNPGPMKQGTGTLGGAGALLG
SFGKRGKGLGVGVGALLGVLGQIGAGMDEODRLAELTSORALETPAGSNVWE
RNPDNNGYGVVTPNKTYRNSTGYCREYQTQTVIGKQKQKAYGNACRQPDGQWQVYV"
60..119
sig_peptide /note="17 kD antigen signal peptide"
120..536
mat_peptide /product="17 kD antigen"
BASE COUNT 179 a 93 c 125 g 142 t
ORIGIN

Query Match 13.3%; Score 64.6; DB 1; Length 539;
Best Local Similarity 49.4%; Pred. No. 1.3e-07;
Matches 197; Conservative 0; Mismatches 199; Indels 3; Gaps 1;

QY 72 CTTTAGTCGTCAGAGTCGCGACTGCGGCTGTGTGGCGGTCTTGTGCGCA 131
Db 134 CATGATAACAGAGTACAGACACTCTTGGCGGTGTGGCGGCCATCTTGGTTC 193
QY 132 GCTGTTTGGTAAAGGTAGTGTGCAATGGCCATTTGGTGCTCTTTTGGGTGG 191
Db 194 TCAATTCGGTAAGGCAAGACAGCTTGTGGAGTAGGTGTAGGTGCTACTTGGAGC 253
QY 192 ATTAATTTGGTCTTAATAATCGGTCAATCGATGATCAGCAGGATAAATAAGCTAAACCA 251
Db 254 AGTTCTTGGTGGCAAAATCGGTGAGGTATGATGAACAGGATAGAAGACTTGCAGAGCT 313
QY 252 GAGTTTGGAAAGGTAAACAGCGGCAAGTGACACGTTGGCGTAATCCAGTACAGCAA 311
```

```
Db 314 TACTCACAGAGCCTTTAGAAC--AGCTCCTAGTGGTAGTAACGTAGATGGCGTAA 370
QY 312 TAGTTATAGTTTACGACAGTGGCTACTTACCAGCGTTACAAATAAGCAAGCGTCGCCA 371
Db 371 TCCGGATAACGGCAATTACGGTTACGTAACACCTAATAAAACTTATAGAAATAGCACTGG 430
QY 372 GCAATATTGTCGAGAAATTTTCAGAAAAGGCGATGATTGCGAGGCGCAGAAAGCAAGATTTA 431
Db 431 TCAATATTGCGCGTACGTAACACTCAACACAGTTGTAATAGGCGGAAACAAACAAAGCATA 490
QY 432 CGGCACGTGCATCGCGCAACCGGATGCTGTTGGCAAGT 470
Db 491 CGGTAATGTCATGCGCCACCTGACGACAATGGCAAGT 529

RESULT 9
LOCUS AE008675 10127 bp DNA linear BCT 14-SEP-2001
DEFINITION Rickettsia conorii Malish 7, section 107 of 114 of the complete
genome.
ACCESSION AE008675 AE006914
VERSION AE008675.1 GI:15620413
KEYWORDS
SOURCE Rickettsia conorii.
ORGANISM Rickettsia conorii
Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
Rickettsiaceae; Rickettsieae; Rickettsia; spotted fever group.
REFERENCE 1 (sites)
AUTHORS Ogata, H., Audic, S., Barbe, V., Artiguenave, F., Fournier, P.E.,
Raoult, D. and Claverie, J.M.
TITLE Selfish DNA in protein-coding genes of Rickettsia
JOURNAL Science 290 (5490), 347-350 (2000)
MEDLINE 20485642
PUBMED 11030655
REFERENCE 2 (sites)
AUTHORS Ogata, H., Audic, S. and Claverie, J.-M.
TITLE Selfish DNA and the origin of genes
JOURNAL Science 291 (5502), 252-253 (2001)
REFERENCE 3 (bases 1 to 10127)
AUTHORS Ogata, H., Audic, S., Renesto-Audiffren, P., Fournier, P.-E., Barbe, V.,
Samson, D., Roux, V., Cossart, P., Weissenbach, J., Claverie, J.-M. and
Raoult, D.
TITLE Mechanisms of evolution in Rickettsia conorii and R. prowazekii
JOURNAL Science 293 (5537), 2093-2098 (2001)
MEDLINE 21442074
PUBMED 11557893
REFERENCE 4 (bases 1 to 10127)
AUTHORS Ogata, H., Audic, S., Renesto-Audiffren, P., Fournier, P.-E., Barbe, V.,
Samson, D., Roux, V., Cossart, P., Weissenbach, J., Claverie, J.-M. and
Raoult, D.
TITLE Direct Submission
JOURNAL Submitted (26-APR-2001) Unite des Rickettsies, Faculte de Medecine,
CNRS UMR 6020, Universite de la Mediterranee, 27 boulevard Jean
Moulin, Marseille Cedex 05 13385, France
COMMENT A public version of R. conorii genome database is accessible at
http://igs-server.cnrs-mrs.fr/. The database intends to provide
updated data. Annotation of the genome is an ongoing task whose
goal is to make the genome sequence more useful. Comments to the
authors are appreciated.
FEATURES
source Location/Qualifiers
1..10127 /organism="Rickettsia conorii"
/db_xref="taxon:781"
111..890
gene /gene="map"
/note="RC1276"
111..890
CDS /gene="map"
/codon_start=1
/transl_table=11
/product="methionine aminopeptidase [EC:3.4.11.18]"
/protein_id="AAL03814.1"
```

```
/db_xref="GI:15620414"
/translation="MTIKHTGKDFIKMRAAGKLAETLDFITDHPVNPVNTSLNGL
CHNFTSHNAVAPLNTKGFPSICTSINHVCIPNDKPKNGDVIINDVTVLIDG
WYDSTMYVGDVAIKPQIOTYIDAMKGIEVRPGAKLGDIGHAIOYAKSHNY
SVRDYTHGIGRVHDEPSILNYSRSTGLTEEGFEFTVEPIMNAGNYDTILSKLD
GWTYTRDKLSAQPEHTIGYTKDGFETFLSPKLDYPPY"
complement(977. .1011)
/feature="REP03, repeated element"
complement(1064. .1083)
/feature="REP02, RS3-like repeat"
1141. .1452
/feature="RC1277"
1141. .1452
/feature="RC1277"
/feature="RC1277"
/codon_start=1
/transl_table=11
/product="similarity to DNA repair protein (RadC)"
/protein_id="AA03815.1"
/db_xref="GI:15620415"
/translation="MLENNLEFSNDEIIDPHYIGHRRKVRKRVAAAGAEYFSDYEL
LEVLFSAPRKDIKPLAKLLDHFDTDLINLDERLLSIKGTNENLYINFALRELI
N"
1457. .1825
/feature="RC1278"
1457. .1825
/feature="RC1278"
/codon_start=1
/transl_table=11
/product="similarity to DNA repair protein (RadC)"
/protein_id="AA03816.1"
/db_xref="GI:15620416"
/translation="MKQKIINKNINASWSDYLVKVMGNMRLEQFRILFLNKKNIL
IADEVLSQGTIDQAAVPEIKRLENEASSLIVHHHPGSGPEPSKANTHMTNKIV
EKQVTINILVNLICSEYK"
1812. .2795
/feature="mray2"
/feature="RC1279"
1812. .2795
/feature="mray2"
/codon_start=1
/transl_table=11
/product="undecaprenyl-phosphate
alpha-N-acetylglucosaminyltransferase [EC:2.4.1.-]"
/protein_id="AA03817.1"
/db_xref="GI:15620417"
/translation="MNINSYSTFISFTAILTYILTKYLTIGLVDIPSNRRSHD
KITPRGGIAIVFVVMIALSGEFTTSNNFVNSIKILPLLAIAIASFLDDLKAVPIL
IRLVHLICAFALFELSQAISVHILYSLIVIALSGEFTINYNEMDGLDGMSCYESIH
LSSTMILICFOPTINNPYFITSVNVIIIGSGCGFLFNHHPAKIFLGDVGSISLGF
LTGLCLLLAULTSNLFTACAIASLYITDAVLTLIRLKNKEKIWOPLHKHFFQKAV
QKGSKHQVWSIIAVCNIFLMIISVTSLSHFVVISIMLALAVITLTQSLLK"
2792. .3115
/feature="RC1280"
2792. .3115
/feature="RC1280"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AA03818.1"
/db_xref="GI:15620418"
/translation="MKLIKEDIDKIVRHIFAKOHPLLPEIMINWKNIVGFNFSPKALP
LKITYTYVKOKINFLFOAENATAELPYQDILIERIKIYLGFAAIHQMNVTFFK
GKSL"
complement(3198. .3227)
/feature="REP02, RS3-like repeat"
3313. .3353
/feature="REP01, repeated element"
complement(3509. .4258)
/feature="RC1281"
complement(3509. .4258)
/feature="RC1281"
/codon_start=1
/transl_table=11
```

```
repeat_region
repeat_region
gene
CDS

/feature="unknown"
/protein_id="AA03819.1"
/db_xref="GI:15620419"
/translation="MKLLLTAAASTALLTSGLSFADCDMNSSVDSSNTSMSSSVEN
QWYLLKAGGVIFENKTPKGADEFKLNKIKSNKSGTGTGEIGAGYIMDNLRDLDLI
GTVASSHLLKSKSTYPDGNFSFVKNKPTIIVSVLLNGYDVFVLDLSPFKVPAGAGVAAFY
KEKIHSDKIKGVYDTFTNGTKNTNFAYQLSLGTSFEVAGVKAELVYSWRDYGKTK
NTTKTINGDKVKFGGTHYKGNHLMAGLRFDN"
4438. .4463
/feature="REP03, repeated element"
complement(4505. .5188)
/feature="RC1282"
complement(4505. .5188)
/feature="RC1282"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AA03820.1"
/db_xref="GI:15620420"
/translation="MLIMKLLLIAAATSMSSSVFAEGMDHEWYLRIDTGAAMFN
EEKDQATGPKTKSNTTPVALGIGYIYSENFADLTGTIIGGLKKSGAATNAPFTTR
TNIASAKHTKITRLLINGYVLTNFMDFDVFAGAGVGSALVKEKITVNGTIGLSNTRK
NRTNISYKLTGTSQAIDAGVKVELAYSWIDDRKTSKNVIYQGTSVPTGCMHQSHN
LTAGIREDI"
complement(5338. .5508)
/feature="REP06, A and T rich repeated element"
complement(5478. .5828)
/feature="fdxA"
/feature="RC1283"
complement(5478. .5828)
/feature="fdxA"
/codon_start=1
/transl_table=11
/product="ferredoxin"
/protein_id="AA03821.1"
/db_xref="GI:15620421"
/translation="WTYVVTDCVKCKYTDCEVCPVDCFYEGREDDFMLVINPDEC
IDCGVCPDCPIGAIKPESGLIEWVERAKDFIENKGWKNITTKTKALPDADKFKDEK
DKFNKYISGVKIAN"
complement(5833. .6525)
/feature="ccmC"
/feature="RC1284"
complement(5833. .6525)
/feature="ccmC"
/codon_start=1
/transl_table=11
/product="heme exporter protein C"
/protein_id="AA03822.1"
/db_xref="GI:15620422"
/translation="MLKLLTPYFNKLSKVIIPIILAISTCVMMIIGLYLALIVSPDY
QQGDFVIMYVHVFPASWMAVGIVFMVCSFSLVWKTTSYLLAVASVYVGFATFLI
SLVTGSLWKPINGTWYWDARLTSMILFLLYLSYIIIVNSADNIRKAQNPSSIAL
IGLINIPIVKFSVNIWYSLHQPASVLRGSPTIHSSMLKPLIIMFISILYFLVLL
RTAILIDKIKNR"
complement(6540. .7046)
/feature="RC1285"
complement(6540. .7046)
/feature="RC1285"
/codon_start=1
/transl_table=11
/product="cytosine deaminase [EC:3.5.4.1]"
/protein_id="AA03823.1"
/db_xref="GI:15620423"
/translation="MNILLDEAFNNFMEQALKQAKIAFDKNEVPVGVAVVDRLHQ
KIIASTHNTEKNNAIYAEITIAINEACNLSKNLNDYDIYVTEPCAMCAAAIAH
SRLARLFYGASDSKHGVVESNLRYFNSSACFRPEIYSGILAEDSGLLMKEFFKIRT
VISSHRMT"
complement(7027. .7947)
/feature="p34"
/feature="RC1286"
complement(7027. .7947)

Query Match 13.3%; Score 64.6; DB 1; Length 10127;
```

| BASE COUNT | 179 a | 92 c | 125 g | 143 t |
|---------------------------|---|--|-------|-------|
| ORIGIN | | | | |
| Query Match | 13.2% Score 64.2; DB 1; Length 539; | | | |
| Best Local Similarity | 49.6% Pred. No. 1.7e-07; | | | |
| Matches 193; Conservative | 0; Mismatches 193; Indels 3; Gaps 1; | | | |
| Qy | 82 | CAAGAAGTCGGAGCTGGGAGCTGGGCGCTGTTGTTGGCGGTGTTGCTGGCCAGCTGTTTGGT | 141 | |
| Db | 144 | CAAGGTACAGGAACACATCTTGGCGGTGCTGGCGGCGCATTTGTTCTCAATTGGT | 203 | |
| Qy | 142 | AAAGGTAGTGTGAGTTGCAATGGCAATGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT | 201 | |
| Db | 204 | AAGGCGCAAGACAGCTGTTGGAGTAGTAGTGTAGGTGTCATTTACTTTGGAGCAGTTCCTTGGT | 263 | |
| Qy | 202 | TCTAAATTCGGTCAATCGATCGATGATCAGCAGGATAAATAAGCTAAACACAGATTTGGAA | 261 | |
| Db | 264 | GGACAATTCGGTGCAGGTATGGATGACAGATAGAGACTTGGCAGAGCTTACCTCACAG | 323 | |
| Qy | 262 | AAGTAAAGCAGGCGCAAGTGACAGTTGGCGTAAATCCAGATACAGGCAATAGTTATAGT | 321 | |
| Db | 324 | AGAGCTTTAGAAAC--AGCTCCTAGTGGTAGTAACTAGAGTGGCGTAAATCCGATAAC | 380 | |
| Qy | 322 | GTTGAGCGAGTGCCTACTTACCAGCGTTACAAATGAAGCAAGCGTCGCGCAGCAATATGT | 381 | |
| Db | 381 | GGCAATTACGGTTACGTAAACACCTTAATAAACTTATAGAAATAGCACTGGTTCATATTC | 440 | |
| Qy | 382 | CGACAATTCAGCAAAAGCGCATGATTCGAGGCGAGAGCAAGAGATTTACGGCACTGCA | 441 | |
| Db | 441 | CGTGAGTACACTAAACAGTTGTAATAGGCGGAAACAAACAAAAGCATACGGTAATGCA | 500 | |
| Qy | 442 | TGCGCGCAACCGGATGCTGTTGGCAAGT | 470 | |
| Db | 501 | TGCGCGCAACCTGACGCAATGCGCAAGT | 529 | |
| RESULT 11 | | | | |
| R1RANTR | R1RANTR 620 bp DNA linear BCT 26-APR-1993 | | | |
| LOCUS | R.rickettsii 17K antigen gene, complete cds. | | | |
| DEFINITION | M16486 | | | |
| ACCESSION | M16486.1 GI:152467 | | | |
| VERSION | antigen. | | | |
| KEYWORDS | R.rickettsii (strain Sheila Smith) DNA, clone pSC2. | | | |
| SOURCE | Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales; Rickettsiaceae; Rickettsieae; Rickettsia; spotted fever group. | | | |
| ORGANISM | 1 (bases 1 to 620) | | | |
| REFERENCE | Anderson,B.E., Regnery,R.L., Carlone,G.M., Tzianabos,T., McDade,J.E., Fu,Z.Y. and Bellini,W.J. | | | |
| AUTHORS | Sequence analysis of the 17-kilodalton-antigen gene from Rickettsia rickettsii | | | |
| TITLE | J. Bacteriol. 169 (6), 2385-2390 (1987) | | | |
| JOURNAL | 87222152 | | | |
| MEDLINE | 3108232 | | | |
| PUBMED | Draft entry and computer-readable sequence for [1] kindly provided by B.E.Anderson, 23-OCT-1987 | | | |
| COMMENT | -35 and -10 regions are located at positions 12-17 and 33-38. | | | |
| FEATURES | Location/Qualifiers | | | |
| source | 1..620 | | | |
| | /organism="Rickettsia rickettsii" | | | |
| | /db_xref="taxon:783" | | | |
| CDS | 70..549 | | | |
| | /note="17kd antigen precursor" | | | |
| | /codon_start=1 | | | |
| | /transl_table=11 | | | |
| | /protein_id="AAA26381.1" | | | |
| | /db_xref="GI:152468" | | | |
| | /translation="MKLLSKIMIIALATSMIQACNGPGGNKNGCTGTLGGAGGALLIG SQFGKQQLVGVGVAGVIGGQIGAGMDQDRRLAELTSSRALETAPSGSNVNEV RPNPDNGYGVTPNRYNSTGYCREYTVVIGGKQKAYGDACRQDPDEQVNVN" | | | |
| | 70..126 | | | |
| | /note="17kd antigen signal peptide" | | | |
| sig_peptide | | | | |

| | | | | | | | | | | |
|-----------------------|---|---|-----|--|--|--|--|--|--|--|
| JOURNAL | Submitted (03-APR-2000) Schulenburg H., Department of Evolutionary Biology, Institut fuer Spezielle Zoologie, Huefferstr. 1, 48149 Muenster, GERMANY | | | | | | | | | |
| FEATURES | Location/Qualifiers | | | | | | | | | |
| source | 1. .434 | | | | | | | | | |
| | /organism="male-killing Rickettsia from Adalia bipunctata" | | | | | | | | | |
| | /specific_host="Adalia bipunctata" | | | | | | | | | |
| | /db_xref="taxon:38028" | | | | | | | | | |
| | /country="United Kingdom: Cambridge" | | | | | | | | | |
| | /notes="isolated from 2spot ladybird" | | | | | | | | | |
| CDS | <1..>434 | | | | | | | | | |
| | /codon_start=1 | | | | | | | | | |
| | /transl_table=11 | | | | | | | | | |
| | /product="17 kDa antigen" | | | | | | | | | |
| | /protein_id="CAB96382.1" | | | | | | | | | |
| | /db_xref="GI:8920295" | | | | | | | | | |
| | /db_xref="SPTREMBL:O9K2N6" | | | | | | | | | |
| | /translation="ALATSLQAQCTPGCMNKOGTCTLLGGAGALLGSQFGKGGKOLVGVGALLGAVLGGTIGAGMDQDRRLAELTTSORALEAAPSGSNVWRPNPDNGNGHYVTPNKYRNSTGYCYREYTGTVVIGGKQKAYGNACKRPDQE" | | | | | | | | | |
| primer_bind | 1. .20 | | | | | | | | | |
| | /note="PCR primer R1" | | | | | | | | | |
| primer_bind | 415. .434 | | | | | | | | | |
| | /note="PCR primer R2" | | | | | | | | | |
| BASE COUNT | 132 a 88 c 114 g 100 t | | | | | | | | | |
| ORIGIN | | | | | | | | | | |
| Query Match | 12.8% Score 62.2; DB 1; Length 434; | | | | | | | | | |
| Best Local Similarity | 50.1%; Pred. No. 6.2e-07; | | | | | | | | | |
| Matches 18; | Conservative 0; Mismatches 178; Indels 3; Gaps 1; | | | | | | | | | |
| Qy | 102 | TGGGGCTGTTTGGCGGTGTGCTGGCCAGCTCTTTGGTAAAGGTAGTGGTCTCAGTTGC | 161 | | | | | | | |
| Db | 75 | TGGCGGTGCCGGCGTGCAATTACTTGGTTCTCAATTCCGCCAAGGTAAGGGCAACTGT | 134 | | | | | | | |
| Qy | 162 | ANTGCCATTGGTGGTGGTGTGTTTGGGTGGATTAAATTGGTTCTAAATCGGTCAATCGAT | 221 | | | | | | | |
| Db | 135 | CGGAGTAGGTGTAGGTGCATTACTTGGAGCAGTTCCTTGGTGACAAATCGGTGCAGGTAT | 194 | | | | | | | |
| Qy | 222 | GGATCAGCAGATAAAATAAAGCTAAACAGAGTTTGAAAAAGCTAAAGCAGGCAAGT | 281 | | | | | | | |
| Db | 195 | GGATGACAGAGATAGAAGACTGTGCTGAACCTCAAGAGAGCTTTAGAACG---AGC | 251 | | | | | | | |
| Qy | 282 | GACACGTTGGCGGTAAATCCAGATACAGGCAATAGTTATAGTTGTGAGCCAGCTGCCTACTTA | 341 | | | | | | | |
| Db | 252 | TCCTAGTGGTAGTAACGTAGAAATGGCGTAAATCCCGATAACGGCAATCATGCTACGTAAC | 311 | | | | | | | |
| Qy | 342 | CCAGCGTTACAAATAAGCAAGAGCTCGCCAGCAATATTTGCGAGAAATTTACGAAAAAGGC | 401 | | | | | | | |
| Db | 312 | ACCTAATAAAACTTTATAGAAATAGCACAGGGCAATATTCCGCTGAGTACACTCAACAGT | 371 | | | | | | | |
| Qy | 402 | GATGATTGCAGGCAGAAAGCAAGAGATTACGGGCACCTGCATGCCGCAACCGGATGTCG | 461 | | | | | | | |
| Db | 372 | TGTAATAGCGGAAACACAAAAAGCATCGGTAAATGATGCCGCCAACCTGACGAACA | 431 | | | | | | | |
| Qy | 462 | TTG 464 | | | | | | | | |
| Db | 432 | ATG 434 | | | | | | | | |
| RESULT 14 | | | | | | | | | | |
| LOCUS | MRI269518 434 bp DNA linear BCT 27-JAN-2001 | | | | | | | | | |
| DEFINITION | Male-killing Rickettsia from Adalia bipunctata partial 17kDa antigen gene (Denmark: Ribe). | | | | | | | | | |
| ACCESSION | AJ269518 | | | | | | | | | |
| VERSION | AJ269518.1 GI:8920296 | | | | | | | | | |
| KEYWORDS | 17 kDa antigen. | | | | | | | | | |
| SOURCE | male-killing Rickettsia from Adalia bipunctata. | | | | | | | | | |
| ORGANISM | Bacteria; Proteobacteria; alpha subdivision; Rickettsiales; Rickettsiaceae; Rickettsieae; Rickettsia. | | | | | | | | | |
| REFERENCE | 1 (bases 1 to 434) | | | | | | | | | |

Search completed: February 22, 2003, 16:09:02
Job time : 1555.32 secs

LOCUS AF195118 546 bp DNA linear BCT 12-JUN-2001
DEFINITION Rickettsia felis 17 kDa genus-common antigen gene, complete cds.
ACCESSION AF195118
VERSION AF195118.1 GI:11066087
KEYWORDS
SOURCE
ORGANISM Rickettsia felis.
Rickettsia felis
Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
Rickettsiaceae; Rickettsieae; Rickettsia; spotted fever group.
1 (bases 1 to 546)
Bouyer,D.H., Stenos,J., Crocquet-Valdes,P., Moron,C.G., Popov,V.L.,
Zavala-Velazquez,J.E., Foil,L.D., Stothard,D.R., Azad,A.F. and
Walker,D.H.
Rickettsia felis: molecular characterization of a new member of the
spotted fever group
Int. J. Syst. Evol. Microbiol. 51 (Pt 2), 339-347 (2001)
21217364
11321078
2 (bases 1 to 546)
Bouyer,D.H., Stenos,J., Crocquet-Valdes,P.A., Foil,L.D. and
Walker,D.H.
Direct Submission
Submitted (14-OCT-1999) Pathology, University of Texas Medical
Branch at Galveston, 301 University Blvd., Galveston, TX
77551-0609, USA
FEATURES
source
1..546
Location/Qualifiers
/organism="Rickettsia felis"
/db_xref="taxon:42862"
65..544
/codon_start=1
/transl_table=11
/product="17 kDa genus-common antigen"
/protein_id="AAG28452.1"
/db_xref="GI:11066088"
/translation="MKLLSKIMIIALASMLQACNPGCMNKQGTGLLGAGGALLG
SQFGKGVGLVGVGALIGAVLGQIGAGMDEQDRRLAELTSORALEATPSGTSVEW
RNPDMNGHYTPNPKYRNSTQYREYTVIIGGKOOKAYGNACRPDGLWQVNV"
BASE COUNT 183 a 96 c 126 g 141 t
ORIGIN
Query Match 12.2%; Score 59.4; DB 1; Length 546;
Best Local Similarity 48.8%; Pred. No. 4.2e-06;
Matches 190; Conservative 0; Mismatches 196; Indels 3; Gaps 1;
QY 82 CAAGAAGTCGGAGCTCGGACTGGGCTGTGTTGGCGGTGTGCTGCCAGCTGTTTGGT 141
DB 149 CAAGGTACGGACACTTCTTGGCGGTGCCGGGTGCATTACTTGGTTCTCAATTGGC 208
QY 142 AAGGTAGTGGTCGATGTCATGGCCATGGTGGTCTGTTTGGTGGATTAATTGGT 201
DB 209 AAGGGCAAAGGACAGCTTCTCGGAGTAGGTGTAGGTGCATTACTTGGAGCAGTCTTGGT 268
QY 202 TCTAAATCGGTCAATCGATGATCAGCAGATATAAATAAGCTAAACCAGAGTTTGGAA 261
DB 269 GGACAAATAGGTGCAGGTATGGATGAGCAGAGTAGAAGACTTGGCTGAACCTCACTTCACAA 328
QY 262 AAGGTAAGAGCAGGCGAAGTGCACGCTTGGCGTAAATCCAGATACAGGCAATAGTTATAGT 321
DB 329 AGAGCTTTAGAACACACCTA--CGGCACTAGCTAGATGGCGTAATCCGGATAAC 385
QY 322 GTTAGCCAGTGGCTACTTACCAGCCTTACATAAGCAAGAGCGTCCGCCAGCAATATTGT 381
DB 386 GGTAAATCATGTTACGTACACCTAATAAACTTATAGAAATAGCAGCTGTCATATATGC 445
QY 382 CGAGAAATTCAGCAAAAGCGATGTTGCGGGCAGACAGAGATTTACGGCACTGCA 441
DB 446 CGTGAGTACACTCAACAGTTGTAATAGCGGAAACACAAAAAGCATACGTAATGCA 505
QY 442 TGCCGGCAACCGGATGGTTCGTTGGCAAGT 470
DB 506 TGCCGGCAACCTGACGACTATGGCAAGT 534

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 22, 2003, 13:08:35 ; Search time 134.862 Seconds

(without alignments)
8065.404 Million cell updates/sec

Title: US-09-677-374-3

Perfect score: 483

Sequence: 1 atgcgtggttgcctgcagg.....aggtgattagcaccgaaaaa 483

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_101002.*

| No. | Score | Query Match | Length | ID | Description |
|-----|-------|-------------|--------|----|-------------|
| 1 | 483 | 100.0 | 483 | 22 | AAF86247 |
| 2 | 483 | 100.0 | 483 | 24 | ABK52402 |
| 3 | 483 | 100.0 | 788 | 22 | AAF86248 |
| 4 | 483 | 100.0 | 788 | 22 | ABK52403 |
| 5 | 295 | 61.1 | 486 | 22 | AAF86246 |
| 6 | 295 | 61.1 | 486 | 24 | ABK52401 |
| 7 | 295 | 61.1 | 489 | 22 | AAH79040 |
| 8 | 118 | 24.4 | 118 | 22 | AAF86254 |
| 9 | 118 | 24.4 | 118 | 24 | ABK52409 |

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|-------------|
| 1 | 483 | 100.0 | 483 | 22 | AAF86247 |
| 2 | 483 | 100.0 | 483 | 24 | ABK52402 |
| 3 | 483 | 100.0 | 788 | 22 | AAF86248 |
| 4 | 483 | 100.0 | 788 | 22 | ABK52403 |
| 5 | 295 | 61.1 | 486 | 22 | AAF86246 |
| 6 | 295 | 61.1 | 486 | 24 | ABK52401 |
| 7 | 295 | 61.1 | 489 | 22 | AAH79040 |
| 8 | 118 | 24.4 | 118 | 22 | AAF86254 |
| 9 | 118 | 24.4 | 118 | 24 | ABK52409 |

| | | | | | | |
|------|-------|------|--------|----|----------|--------------------|
| 10 | 108.4 | 22.4 | 110 | 22 | AAF86252 | PCR primer #2 used |
| 11 | 108.4 | 22.4 | 110 | 24 | ABK52407 | Escherichia coli c |
| c 12 | 102 | 21.1 | 102 | 22 | AAF86255 | PCR primer #5 used |
| c 13 | 102 | 21.1 | 102 | 24 | ABK52410 | Escherichia coli c |
| 14 | 92.4 | 19.1 | 94 | 22 | AAF86253 | PCR primer #3 used |
| 15 | 92.4 | 19.1 | 94 | 24 | ABK52408 | Escherichia coli c |
| 16 | 74 | 15.3 | 111 | 22 | AAF86251 | PCR primer #1 used |
| 17 | 74 | 15.3 | 111 | 24 | ABK52406 | Escherichia coli c |
| c 18 | 72.4 | 15.0 | 110 | 22 | AAF86256 | PCR primer #6 used |
| c 19 | 72.4 | 15.0 | 110 | 24 | ABK52411 | Escherichia coli c |
| c 20 | 39.4 | 8.2 | 2292 | 21 | AAC45883 | Arabidopsis thalia |
| c 21 | 36.6 | 7.6 | 303 | 16 | AAT18913 | DNA encoding spide |
| c 22 | 36.6 | 7.6 | 303 | 16 | AAT18912 | Erythrobacter long |
| c 23 | 36 | 7.5 | 1305 | 17 | AAT31799 | DNA encoding spide |
| c 24 | 36 | 7.5 | 3251 | 23 | ABL03846 | Drosophila melanog |
| c 25 | 36 | 7.5 | 7791 | 22 | AAI37335 | Human musculocele |
| c 26 | 35.6 | 7.4 | 292 | 21 | AAA45539 | Human secreted exp |
| c 27 | 35.6 | 7.4 | 2481 | 20 | AAK90448 | Human secreted pro |
| c 28 | 35.6 | 7.4 | 2481 | 22 | AAK59277 | Human cDNA encodin |
| c 29 | 35.6 | 7.4 | 2481 | 24 | ABA90946 | Human polynucleoti |
| c 30 | 35.4 | 7.3 | 571 | 20 | AAK61355 | DNA encoding a hum |
| c 31 | 35.4 | 7.3 | 1290 | 19 | AAV36619 | Nucleotide sequenc |
| c 32 | 35.4 | 7.3 | 27541 | 22 | AAD17185 | Streptomyces nouns |
| c 33 | 35.4 | 7.3 | 125401 | 22 | AAI17186 | Streptomyces nouns |
| c 34 | 35.2 | 7.3 | 1479 | 21 | AAA51609 | HIV synthetic gag |
| c 35 | 35.2 | 7.3 | 1479 | 21 | AAA51609 | Synthetic gag poly |
| c 36 | 35 | 7.2 | 77536 | 21 | AAA14651 | Nucleotide sequenc |
| c 37 | 34.8 | 7.2 | 9407 | 23 | AAK59606 | Propionibacterium |
| c 38 | 34.6 | 7.2 | 77536 | 21 | AAA14651 | Nucleotide sequenc |
| c 39 | 34.2 | 7.1 | 6273 | 23 | ABU15731 | Drosophila melanog |
| c 40 | 34.2 | 7.1 | 10855 | 23 | ABU15730 | Drosophila melanog |
| c 41 | 34 | 7.0 | 2652 | 20 | AAK23386 | Human Delta7-ster |
| c 42 | 34 | 7.0 | 2652 | 20 | AAK23387 | Human Delta7-ster |
| c 43 | 33.8 | 7.0 | 810 | 22 | AAH66598 | C glutamicum codin |
| c 44 | 33.8 | 7.0 | 933 | 22 | AAF68080 | Corynebacterium gl |
| c 45 | 33.8 | 7.0 | 1062 | 24 | ABQ90125 | M. capsulatus gene |

ALIGNMENTS

RESULT 1

AAF86247 standard; DNA; 483 BP.

ID AAF86247

XX AC

XX AAF86247;

XX AC

DT 11-JUL-2001 (first entry)

XX DNA sequence of E. coli optimised ospA gene 17E2.

DE Poikilothermic fish; Piscirickettsia salmonis; rickettsial pathogen;

XX KW vaccine; OspA; salmonid rickettsial septicaemia; rickettsial disease;

XX KW SRS; 17E2; ds.

XX OS

OS Piscirickettsia salmonis.

OS Synthetic.

XX Key

XX Location/Qualifiers

FT CDS

FT 1..483

FT /*tag= a

FT /partial

FT /product= "OspA"

FT /note= "Genus specific 17kDa antigen, the sequence does not include a stop codon"

XX CA2281913-AL.

PN 17-MAR-2001.

PD 17-SEP-1999;

XX 99CA-2281913.

XX 17-SEP-1999;

XX 99CA-2281913.

PR 17-SEP-1999;

XX 99CA-2281913.

XX (KAYW/) KAY W W.
PA (BURI/) BURIAN J.
PA (KUYZ/) KUYZIK M A.
XX
PI Kay WW, Burian J, Kuzyk MA;
XX
DR WPI; 2001-316844/34.
DR P-PSDB; AAB81127.
XX
XX Method for protecting poikilothermic fish against salmonid rickettsial
PT septicemia and other rickettsial diseases comprises administering a
PT vaccine containing the Ospa protein of *Piscirickettsia salmonis* -
XX
PS Example 3; Fig 4C; 35pp; English.
XX
CC This invention relates to a method for the protection against infection
CC of a poikilothermic fish by the bacterial pathogen, *Piscirickettsia*
CC *salmonis*. The method comprises administering an immunogenic amount of a
CC *P. salmonis* specific antigen termed Ospa, or an immunogenic fragment of
CC Ospa in the form of a vaccine. The method is used for protecting animals,
CC particularly poikilothermic fish, against the bacterial pathogen
CC *P. salmonis*. The method is also useful for protecting against salmonid
CC rickettsial septicemia (SR) and other rickettsial diseases. The present
CC sequence represents DNA which has been optimised for expression in
CC *Escherichia coli* to encode the *P. salmonis* Ospa protein. An Ospa protein
CC with an N-terminal fusion partner is used in a vaccine to create an
CC anti-Ospa antibody response.
XX
SQ Sequence 483 BP; 108 A; 127 C; 148 G; 100 T; 0 other;

Query Match 100.0%; Score 483; DB 22; Length 483;
Best Local Similarity 100.0%; Pred. No. 5.1e-136;
Matches 483; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGCGTGGTTGCTGCGAGGCGAGCTCTCTGATCATATCTCTGTTTCTCTGGTGGTTGC 60
Db 1 ATGCGTGGTTGCTGCGAGGCGAGCTCTCTGATCATATCTCTGTTTCTCTGGTGGTTGC 60
QY 61 GCCCAGAACTTACGCGCGCAGGAAGTTGGCGGCGCCAGCGGTGCGGTGCGGTGTT 120
Db 61 GCCCAGAACTTACGCGCGCAGGAAGTTGGCGGCGCCAGCGGTGCGGTGCGGTGTT 120
QY 121 GCGGCGCAGCTGTCGGTAAAGCTCTGCTGCTGTGTCGATGCCATCGCGGTGCGGTT 180
Db 121 GCGGCGCAGCTGTCGGTAAAGCTCTGCTGCTGTGTCGATGCCATCGCGGTGCGGTT 180
QY 181 CTGGGCGGTCTGATGGCTCTAAATCGCTCAGCATGCGCAGCAGCATTAATCAAA 240
Db 181 CTGGGCGGTCTGATGGCTCTAAATCGCTCAGCATGCGCAGCAGCATTAATCAAA 240
QY 241 CTGAACAGCTCTCTGAAAGAGTGAAGCGCGCGAGTTACTCTGTTGGCGTAATCCGGAC 300
Db 241 CTGAACAGCTCTCTGAAAGAGTGAAGCGCGCGAGTTACTCTGTTGGCGTAATCCGGAC 300
QY 301 ACCGTTAAGCTACTCTGTGAACCGGTTTCGACCTACCGGTTTACAAACAGGAA 360
Db 301 ACCGTTAAGCTACTCTGTGAACCGGTTTCGACCTACCGGTTTACAAACAGGAA 360
QY 361 GCGCGTACAGCTACTGCGCGCAATTCAGCAGAACGATGATCCAGGTGAGAACAG 420
Db 361 GCGCGTACAGCTACTGCGCGCAATTCAGCAGAACGATGATCCAGGTGAGAACAG 420
QY 421 GAAATCTAGGCGCAGCGTGCCTCAGCGGATGGCGGTGAGGTGATGACCCGAA 480
Db 421 GAAATCTAGGCGCAGCGTGCCTCAGCGGATGGCGGTGAGGTGATGACCCGAA 480
QY 481 AAA 483
Db 481 AAA 483

RESULT 2

ABK52402
ID ABK52402 standard; DNA; 483 BP.
XX
AC ABK52402;
XX
DT 12-AUG-2002 (first entry)
XX
DE DNA encoding *Escherichia coli* codon optimised Ospa, 17e2.
XX
KW Outer surface lipoprotein; Ospa; antibacterial; immunosuppressive;
KW vaccine; poikilothermic fish; fin-fish; Rickettsial septicemia;
KW Rickettsial disease; gene; ds; 17e2.
XX
OS *Piscirickettsia salmonis*.
OS Synthetic.
XX
FH Key Location/Qualifiers
FT CDS 1..483
FT /*tag= a
FT /product= "17e2"
FT /note= "Escherichia coli codon optimised Ospa"
FT /partial
FT /note= "No stop codon given"
XX
PN CA2339327-Al.
XX
PD 15-MAR-2002.
XX
PF 19-MAR-2001; 2001CA-2339327.
XX
PR 15-SEP-2000; 2000US-0677374.
XX
PA (THOR/) THORNTON J C.
PA (KAYW/) KAY W W.
PA (BURI/) BURIAN J.
PA (KUYZ/) KUYZIK M A.
XX
PI Thornton JC, Kay WW, Burian J, Kuzyk MA;
DR WPI; 2002-455221/49.
DR P-PSDB; AAD97868.
XX
PT Inducing immunity in fin fish to Rickettsial septicemia, comprises
PT administration of an outer surface lipoprotein (Ospa) of a bacterial
PT strain, as a vaccine -
XX
PS Claim 16; Fig 4; 55pp; English.
XX
CC The invention describes a method of protecting a poikilothermic fish
CC against infection by the bacterial pathogen *Piscirickettsia salmonis*
CC comprising administering either intraperitoneally, by immersion or
CC orally, an immunogenic amount of principal antigen, the Ospa (outer
CC surface lipoprotein), its variants, non-lipidated form or antigenic
CC peptides derived or synthesized with or without an adjuvant. The new
CC method is used to provide an outer surface lipoprotein (Ospa) of
CC bacterial strain *Piscirickettsia salmonis* as a vaccine to induce immunity
CC in fin-fish against Rickettsial septicemia and other related
CC Rickettsial diseases caused by either a virus, bacteria or parasite.
CC This sequence encodes the *Escherichia coli* codon optimised outer surface
CC lipoprotein Ospa (17e2) used in the creation of the vaccine described in
CC the invention.
XX
SQ Sequence 483 BP; 108 A; 127 C; 148 G; 100 T; 0 other;

Query Match 100.0%; Score 483; DB 24; Length 483;
Best Local Similarity 100.0%; Pred. No. 5.1e-136;
Matches 483; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ATGCGTGGTTGCTGCGAGGCGAGCTCTCTGATCATATCTCTGTTTCTCTGGTGGTTGC 60
Db 1 ATGCGTGGTTGCTGCGAGGCGAGCTCTCTGATCATATCTCTGTTTCTCTGGTGGTTGC 60
QY 61 GCCCAGAACTTACGCGCGCAGGAAGTTGGCGGCGCCAGCGGTGCGGTGTTGCGGTGTT 120


```

Db 61 GCCCAGAACTTCAGCGCCAGAGAAAGTTGGCGGCGCCACCGGTGCGGTGGCGGTGTT 120
QY 121 GCCGCGCAGCTGTTCCGTTAAAGGCTCTGTCGTGTGTCGATGCCCATCGCGGTGCGGTT 180
Db 121 GCCGCGCAGCTGTTCCGTTAAAGGCTCTGTCGTGTGTCGATGCCCATCGCGGTGCGGTT 180
QY 181 CTGGGCGGTCTGATGGCTCTAAATCGTTCAGACATGGACAGCAGGATAAAATCAAA 240
Db 181 CTGGGCGGTCTGATGGCTCTAAATCGTTCAGACATGGACAGCAGGATAAAATCAAA 240
QY 241 CTGAACCACTCTCTGAAAAAGTGAAGCGCGCCAGGTTACTCGTTGGCGTAAATCCGGAC 300
Db 241 CTGAACCACTCTCTGAAAAAGTGAAGCGCGCCAGGTTACTCGTTGGCGTAAATCCGGAC 300
QY 301 ACCGGTAAACAGTACTCTCTGGAACCGGTTTCGACACCTACCGGTTTACAAACAACAGGAA 360
Db 301 ACCGGTAAACAGTACTCTCTGGAACCGGTTTCGACACCTACCGGTTTACAAACAACAGGAA 360
QY 361 CGCGCTCAGCAGTACTGCGCGGAAATTTACAGCAAAAGCCATGATCGCAGGTTCAGAAACAG 420
Db 361 CGCGCTCAGCAGTACTGCGCGGAAATTTACAGCAAAAGCCATGATCGCAGGTTCAGAAACAG 420
QY 421 GAAATCTACGCGACCGCTGCGCTCAGCGGATGGCGCGTGGCAGGTGATTAGCACCAGAA 480
Db 421 GAAATCTACGCGACCGCTGCGCTCAGCGGATGGCGCGTGGCAGGTGATTAGCACCAGAA 480
QY 481 AAA 483
Db 481 AAA 483

RESULT 3
ID AAF86248
ID AAF86248 standard; DNA; 768 BP.
XX AAF86248;
AC AAF86248;
XX 11-JUL-2001 (first entry)
XX DNA sequence of cl7e2 ospA construct with N-terminal fusion partner.
KW Poikilothermic fish; Piscirickettsia salmonis; rickettsial pathogen;
KW vaccine; OspA; salmonid rickettsial septicaemia; rickettsial disease;
KW SR5; 17E2; fusion construct; ds.
XX Piscirickettsia salmonis.
OS Synthetic.
XX Key Location/Qualifiers
FH CDS 1..768
FT /*tag= a
FT /partial
FT /product= "cl7E2 OspA with N-terminal fusion partner"
FT /note= "No stop codon is given"
FT misc_feature 1..285
FT /*tag= b
FT /note= "DNA encoding undefined N-terminal fusion partner"
FT misc_feature 286..768
FT /*tag= c
FT /note= "Optimised OspA construct cl7E2"
XX CA2281913-A1.
XX 17-MAR-2001.
XX 17-SEP-1999; 99CA-2281913.
XX 17-SEP-1999; 99CA-2281913.
XX (KAYW/) KAY W W.
XX (BURI/) BURIAN J.
XX (KUZV/) KUZIK M A.
```

```

XX Kay WW, Burian J, Kuzik MA;
XX WPI; 2001-316844/34.
DR P-PSDB; AAB81128.
XX Method for protecting poikilothermic fish against salmonid rickettsial
PT septicaemia and other rickettsial diseases comprises administering a
PT vaccine containing the ospA protein of Piscirickettsia salmonis -
XX Example 4; Fig 5; 35pp; English.
XX This invention relates to a method for the protection against infection
CC of a poikilothermic fish by the bacterial pathogen, Piscirickettsia
CC salmonis. The method comprises administering an immunogenic fragment of a
CC P. salmonis specific antigen termed OspA, or an immunogenic fragment of
CC OspA in the form of a vaccine. The method is used for protecting animals,
CC particularly poikilothermic fish, against the bacterial pathogen
CC P. salmonis. The method is also useful for protecting against salmonid
CC rickettsial septicaemia (SR5) and other rickettsial diseases. The present
CC sequence represents P. salmonis OspA DNA termed cl7E2 optimised for
CC expression in Escherichia coli fused to DNA encoding an undefined
CC N-terminal fusion partner. The protein encoded by this fusion construct
CC is used in a vaccine to create an anti-OspA antibody response.
XX Sequence 768 BP; 220 A; 183 C; 190 G; 175 T; 0 other;
SQ Query Match 100.0%; Score 483; DB 22; Length 768;
Best Local Similarity 100.0%; Pred. No. 6.2e-136;
Matches 483; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ATCGGTGGTTCCTGAGGCGACGCTCTGATCATTCATCTCTTTTCCTGGGTGTC 60
Db 286 ATCGGTGGTTCCTGAGGCGACGCTCTGATCATTCATCTCTTTTCCTGGGTGTC 345
QY 61 GCCCAGAACTTCAGCGCCAGAGAAAGTTGGCGGCGCCACCGGTGCGGTGTT 120
Db 346 GCCCAGAACTTCAGCGCCAGAGAAAGTTGGCGGCGCCACCGGTGCGGTGTT 405
QY 121 GCCGCGCAGCTGTTCCGTTAAAGGCTCTGTCGTGTGTCGATGCCCATCGCGGT 180
Db 406 GCCGCGCAGCTGTTCCGTTAAAGGCTCTGTCGTGTGTCGATGCCCATCGCGGT 465
QY 181 CTGGGCGGTCTGATGGCTCTAAATCGTTCAGACATGGACAGCAGGATAAAATCAAA 240
Db 466 CTGGGCGGTCTGATGGCTCTAAATCGTTCAGACATGGACAGCAGGATAAAATCAAA 525
QY 241 CTGAACCACTCTCTGAAAAAGTGAAGCGCGCCAGGTTACTCGTTGGCGTAAATCCGGAC 300
Db 526 CTGAACCACTCTCTGAAAAAGTGAAGCGCGCCAGGTTACTCGTTGGCGTAAATCCGGAC 585
QY 301 ACCGGTAAACAGTACTCTCTGGAACCGGTTTCGACACCTACCGGTTTACAAACAACAGGAA 360
Db 586 ACCGGTAAACAGTACTCTCTGGAACCGGTTTCGACACCTACCGGTTTACAAACAACAGGAA 645
QY 361 CGCGCTCAGCAGTACTGCGCGGAAATTTACAGCAAAAGCCATGATCGCAGGTTCAGAAACAG 420
Db 646 CGCGCTCAGCAGTACTGCGCGGAAATTTACAGCAAAAGCCATGATCGCAGGTTCAGAAACAG 705
QY 421 GAAATCTACGCGACCGCTGCGCTCAGCGGATGGCGCGTGGCAGGTGATTAGCACCAGAA 480
Db 706 GAAATCTACGCGACCGCTGCGCTCAGCGGATGGCGCGTGGCAGGTGATTAGCACCAGAA 765
QY 481 AAA 483
Db 766 AAA 768

RESULT 4
ABK52403
ID ABK52403 standard; DNA; 768 BP.
XX ABK52403;
AC ABK52403;
```


DR WPI: 2001-316844/34.
 PT P-PSDB; AAB81126.
 XX
 PT Method for protecting poikilothermic fish against salmonid rickettsial
 PT septicemia and other rickettsial diseases comprises administering a
 PT vaccine containing the OSPA protein of *Piscirickettsia salmonis*.
 XX
 PS Disclosure; Fig 2B; 35pp; English.
 XX
 CC This invention relates to a method for the protection against infection
 CC of a poikilothermic fish by the bacterial pathogen, *Piscirickettsia*
 CC salmonis. The method comprises administering an immunogenic amount of a
 CC P. salmonis specific antigen termed OSPA, or an immunogenic fragment of
 CC OSPA in the form of a vaccine. The method is used for protecting animals,
 CC particularly poikilothermic fish, against the bacterial pathogen
 CC P. salmonis. The method is also useful for protecting against salmonid
 CC rickettsial septicemia (SRS) and other rickettsial diseases. The present
 CC sequence represents DNA encoding the P. salmonis OSPA protein. An OSPA
 CC protein with an N-terminal fusion partner is used in a vaccine to create
 CC an anti-OSPA antibody response.
 XX
 SQ Sequence 486 BP; 137 A; 79 C; 144 G; 126 T; 0 other;
 Query Match 61.1%; Score 295; DB 22; Length 486;
 Best Local Similarity 76.0%; Pred. No. 3.7e-79;
 Matches 364; Conservative 0; Mismatches 115; Indels 0; Gaps 0;
 QY 5 GTGGTTGCCCTGACGGCAGCTCTCTGATCATTTATCTCTGTTTCTGCTGGTGGTGGCC 64
 DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 QY 8 GAGGATGTTTGAAGGAGTAGTAGTCTAATTTATCATGCTGTGTTTATGTTGGCTGCC 67
 DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 QY 65 AGAAGCTTACGCCAGGAGTAGTGGCGCGCCACCGTGGTGGTGGTGGTGGTGGTGG 124
 DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 QY 68 AGAAGCTTACGCCAGGAGTAGTGGCGCGCCACCGTGGTGGTGGTGGTGGTGGTGG 127
 DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 QY 125 GCCAGCTGTTTCGTAAGGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 184
 DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 QY 128 GCCAGCTGTTTCGTAAGGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 187
 DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 QY 185 GCGGCTGATGGCTCTAAAATCGGTGAGAGTGGACGAGCAGGATAAATCAACTGA 244
 DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 QY 188 GTGGATTAATGTTGTTCTAAAATCGGTGAGAGTGGACGAGCAGGATAAATCAACTGA 247
 DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 QY 245 ACCAGCTCTCGGAAAGTGAAGCCCGCCAGGTTACTCGTGGCGTGAATCCGACACCG 304
 DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 QY 248 ACCAGGTTTGGAAAGGTAAGCCCGCCAGGTTACTCGTGGCGTGAATCCGACACCG 307
 DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 QY 305 GTACAGCTACTCTGTGGAACCGTTTGGCAGCTTACAGCGTTTACAAACAGGACGCC 364
 DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 QY 308 GCAATAGTTATAGTTGACCCAGTGGTACTTACCAGCGTTTACATGACAGAGCGTC 367
 DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 QY 365 GTCAGCAGTACTGCCGGAATTTTCAGCAGAAAGCCATGATCGCAGGTTCAGAAACAGGAAA 424
 DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 QY 368 GCCAGCAATATTGTCGAGATTTTCAGCAAAAGCGGATGATTCGAGGCGCAGAACGAGA 427
 DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 QY 425 TCTACGCCACCGTGGCCCTCAGCCGATGGCGCTGGCGAGGTGATTTAGCACCGAAAAA 483
 DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 QY 428 TTTACGGCACTGATGCCGCAACCGGATGTCGTTGGCAAGTCATTTTCAACAGAAAAA 486
 DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 RESULT 6
 ABK52401
 ID ABK52401 standard; DNA; 486 BP.
 XX
 AC ABK52401;
 XX
 DT 12-AUG-2002 (first entry)
 XX
 DE DNA encoding outer surface lipoprotein OSPA.
 XX
 KW Outer surface lipoprotein; OSPA; antibacterial; immunosuppressive;
 KW vaccine; poikilothermic fish; fin-fish; Rickettsial septicemia;
 KW Rickettsial disease; gene; ds.

XX *Piscirickettsia salmonis*.
 OS
 XX
 FH Key Location/Qualifiers
 FT CDS 1..486
 FT /tag- a
 FT /product= "ospa"
 FT /note= "Outer surface lipoprotein"
 FT /partial= "No stop codon given"
 XX
 XX CA2339327-A1.
 XX
 XX 15-MAR-2002.
 XX
 XX 19-MAR-2001; 2001CA-2339327.
 XX
 XX 15-SEP-2000; 2000US-0677374.
 XX
 XX (THOR/) THORNTON J C.
 XX (KAYW/) KAY W W.
 XX (BURI/) BURIAN J.
 XX (KUYZ/) KUYZ M A.
 XX
 XX Thornton JC, Kay WW, Burian J, Kuzyk MA;
 XX WPI: 2002-455221/49.
 XX P-PSDB; AAU97867.
 DR
 DR Inducing immunity in fin fish to Rickettsial septicemia, comprises
 PT administration of an outer surface lipoprotein (OSPA) of a bacterial
 PT strain, as a vaccine.
 XX
 XX Claim 19; Fig 2; 55pp; English.
 PS
 XX The invention describes a method of protecting a poikilothermic fish
 CC against infection by the bacterial pathogen *Piscirickettsia salmonis*
 CC comprising administering either intraperitoneally, by immersion or
 CC orally, an immunogenic amount of principal antigen, the OSPA (outer
 CC surface lipoprotein), its variants, non-lipidated form or antigenic
 CC peptides derived or synthesized with or without an adjuvant. The new
 CC method is used to provide an outer surface lipoprotein (OSPA) of
 CC bacterial strain *Piscirickettsia salmonis* as a vaccine to induce immunity
 CC in fin-fish against Rickettsial septicemia and other related
 CC Rickettsial diseases caused by either a virus, bacteria or parasite.
 CC This sequence encodes the *Piscirickettsia salmonis* outer surface
 CC lipoprotein, OSPA, used in the creation of the vaccine described in the
 CC invention.
 XX
 SQ Sequence 486 BP; 137 A; 79 C; 144 G; 126 T; 0 other;
 Query Match 61.1%; Score 295; DB 24; Length 486;
 Best Local Similarity 76.0%; Pred. No. 3.7e-79;
 Matches 364; Conservative 0; Mismatches 115; Indels 0; Gaps 0;
 QY 5 GTGGTTGCCCTGACGGCAGCTCTCTGATCATTTATCTCTGTTTCTGCTGGTGGTGGCC 64
 DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 DB 8 GAGGATGTTTGAAGGAGTAGTAGTCTAATTTATCATGCTGTGTTTATGTTGGCTGCC 67
 QY 65 AGAAGCTTACGCCAGGAGTAGTGGCGCGCCACCGTGGTGGTGGTGGTGGTGGTGG 124
 DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 DB 68 AGAAGCTTACGCCAGGAGTAGTGGCGCGCCACCGTGGTGGTGGTGGTGGTGGTGG 127
 QY 125 GCCAGCTGTTTCGTAAGGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 184
 DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 DB 128 GCCAGCTGTTTCGTAAGGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 187
 QY 185 GCGGCTGATGGCTCTAAAATCGGTGAGAGTGGACGAGCAGGATAAATCAACTGA 244
 DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 DB 188 GTGGATTAATGTTGTTCTAAAATCGGTGAGAGTGGACGAGCAGGATAAATCAACTGA 247
 QY 245 ACCAGCTCTCGGAAAGTGAAGCCCGCCAGGTTACTCGTGGCGTGAATCCGACACCG 304
 DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 DB 248 ACCAGGTTTGGAAAGGTAAGCCCGCCAGGTTACTCGTGGCGTGAATCCGACACAG 307

CC Ospa in the form of a vaccine. The method is used for protecting animals,
CC particularly poikilothermic fish, against the bacterial pathogen
CC P. salmonis. The method is also useful for protecting against salmonid
CC rickettsial septicaemia (SRS) and other rickettsial diseases. The present
CC sequence represents a PCR primer used in the cloning and optimisation of
CC the P. salmonis Ospa gene. The Ospa gene is used in the method of the
CC invention.

XX
SQ Sequence 118 BP; 21 A; 32 C; 28 G; 37 T; 0 other;
Query Match 24.4%; Score 118; DB 22; Length 118;
Best Local Similarity 100.0%; Pred. No. 7.1e-26;
Matches 118; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 208 GGTGAGGATGACGACGAGGATAAATCAAACTGAACAGTCTCTGGAAGTGAAG 267
Db 118 GGTGAGGATGACGACGAGGATAAATCAAACTGAACAGTCTCTGGAAGTGAAG 59
QY 268 GCCGCCAGGTTACTCTGTTGGCGTAATCCGGACACCGGTAACTGCTGTGGAAAC 325
Db 58 GCCGCCAGGTTACTCTGTTGGCGTAATCCGGACACCGGTAACTGCTGTGGAAAC 1

RESULT 9
ABK52409/C
ID ABK52409 standard; DNA; 118 BP.
XX
AC ABK52409;
XX
DT 12-AUG-2002 (first entry)
XX
DE Escherichia coli codon optimised Ospa 17e2, oligonucleotide #4.
XX
KW Outer surface lipoprotein; Ospa; antibacterial; immunosuppressive;
KW vaccine; poikilothermic fish; fin-fish; Rickettsial septicaemia;
KW Rickettsial disease; 17e2; ss.
XX
OS Synthetic.
XX
CA2339327-A1.
XX
15-MAR-2002.
XX
PF 19-MAR-2001; 2001CA-2339327.
XX
PR 15-SEP-2000; 2000US-0677374.
XX
PA (THOR/) THORNTON J C.
PA (KAYW/) KAY W W.
PA (BURI/) BURIAN J.
PA (KUZU/) KUZU M A.
XX
PI Thornton JC, Kay WW, Burian J, Kuzyk MA;
XX WPI; 2002-455221/49.
XX
PT Inducing immunity in fin fish to Rickettsial septicaemia, comprises
PT administration of an outer surface lipoprotein (Ospa) of a bacterial
PT strain, as a vaccine -
XX
PS Example 3; Fig 4; 55pp; English.
XX
CC The invention describes a method of protecting a poikilothermic fish
CC against infection by the bacterial pathogen Piscirickettsia salmonis
CC comprising administering either intraperitoneally, by immersion or
CC orally, an immunogenic amount of principal antigen, the Ospa (outer
CC surface lipoprotein), its variants, non-lipidated form or antigenic
CC peptides derived or synthesized with or without an adjuvant. The new
CC method is used to provide an outer surface lipoprotein (Ospa) of
CC bacterial strain Piscirickettsia salmonis as a vaccine to induce immunity
CC in fin-fish against Rickettsial septicaemia and other related
CC Rickettsial diseases caused by either a virus, bacteria or parasite.
CC This sequence represents an oligonucleotide used to construct an

CC Escherichia coli codon optimised outer surface lipoprotein Ospa (17e2)
XX used in the creation of the vaccine described in the invention.
SQ Sequence 118 BP; 21 A; 32 C; 28 G; 37 T; 0 other;
Query Match 24.4%; Score 118; DB 24; Length 118;
Best Local Similarity 100.0%; Pred. No. 7.1e-26;
Matches 118; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 208 GGTGAGGATGACGACGAGGATAAATCAAACTGAACAGTCTCTGGAAGTGAAG 267
Db 118 GGTGAGGATGACGACGAGGATAAATCAAACTGAACAGTCTCTGGAAGTGAAG 59
QY 268 GCCGCCAGGTTACTCTGTTGGCGTAATCCGGACACCGGTAACTGCTGTGGAAAC 325
Db 58 GCCGCCAGGTTACTCTGTTGGCGTAATCCGGACACCGGTAACTGCTGTGGAAAC 1

RESULT 10
AAF86252
ID AAF86252 standard; DNA; 110 BP.
XX
AC AAF86252;
XX
DT 11-JUL-2001 (first entry)
XX
DE PCR primer #2 used in cloning an optimisation of Ospa gene.
XX
KW Poikilothermic fish; Piscirickettsia salmonis; rickettsial pathogen;
KW vaccine; Ospa; salmonid rickettsial septicaemia; rickettsial disease;
KW SRS; 17E2; fusion construct; PCR primer; ss.
XX
OS Piscirickettsia salmonis.
XX
CA2281913-A1.
XX
PD 17-MAR-2001.
XX
PF 17-SEP-1999; 99CA-2281913.
XX
PR 17-SEP-1999; 99CA-2281913.
XX
PA (KAYW/) KAY W W.
PA (BURI/) BURIAN J.
PA (KUZU/) KUZU M A.
XX
PI Kay WW, Burian J, Kuzyk MA;
XX WPI; 2001-316844/34.
XX
PT Method for protecting poikilothermic fish against salmonid rickettsial
PT septicaemia and other rickettsial diseases comprises administering a
PT vaccine containing the Ospa protein of Piscirickettsia salmonis -
XX
PS Example 3; Fig 4B; 35pp; English.
XX
CC This invention relates to a method for the protection against infection
CC of a poikilothermic fish by the bacterial pathogen, Piscirickettsia
CC salmonis. The method comprises administering an immunogenic amount of a
CC P. salmonis specific antigen termed Ospa, or an immunogenic fragment of
CC Ospa in the form of a vaccine. The method is used for protecting animals,
CC particularly poikilothermic fish, against the bacterial pathogen
CC P. salmonis. The method is also useful for protecting against salmonid
CC rickettsial septicaemia (SRS) and other rickettsial diseases. The present
CC sequence represents a PCR primer used in the cloning and optimisation of
CC the P. salmonis Ospa gene. The Ospa gene is used in the method of the
CC invention.

XX
SQ Sequence 110 BP; 13 A; 28 C; 45 G; 24 T; 0 other;
Query Match 22.4%; Score 108.4; DB 22; Length 110;
Best Local Similarity 99.1%; Pred. No. 5.5e-23;
Matches 109; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 53 TGGGTTGGCCCGCCAGAACTTCAGCCGCCAGGAAGTTGGCGCGCCACCGGTGCGGTGTGG 112
|||||
Db 1 TGGGTTGGCCCGCCAGAACTTCAGCCGCCAGGAAGTTGGCGCGCCACCGGTGCGGTGTGG 60

QY 113 GCGGTTGGCCCGCCAGCTGTTCCGTTAAAGGCTCTGGTCTGTCTGTCGATG 162
|||||
Db 61 GCGGTTGGCCCGCCAGCTGTTCCGTTAAAGGCTCTGGTCTGTCTGTCGATG 110

RESULT 11
ABK52407
ID ABK52407 standard; DNA; 110 BP.
XX
AC ABK52407;
XX
XX 12-AUG-2002 (first entry)
XX
XX Escherichia coli codon optimised OspA 17e2, oligonucleotide #2.
XX
KW Outer surface lipoprotein; OspA; antibacterial; immunosuppressive;
KW vaccine; poikilothermic fish; fin-fish; Rickettsial septicaemia;
KW Rickettsial disease; 17e2; ss.
XX
OS Synthetic.
XX
PN CA2339327-Al.
XX
PD 15-MAR-2002.
XX
PF 19-MAR-2001; 2001CA-2339327.
XX
XX 15-SEP-2000; 2000US-0677374.
XX
XX (THOR/) THORNTON J C.
PA (KAYW/) KAY W W.
PA (BURL/) BURIAN J.
PA (KUZY/) KUZIK M A.
XX
PI Thornton JC, Kay WW, Burian J, Kuzyk MA;
XX
DR WPI; 2002-455221/49.
XX
XX Inducing immunity in fin fish to Rickettsial septicaemia, comprises
PT administration of an outer surface lipoprotein (OspA) of a bacterial
PT strain, as a vaccine -
XX
XX Example 3; Fig 4; 55pp; English.
XX
XX The invention describes a method of protecting a poikilothermic fish
CC against infection by the bacterial pathogen Piscirickettsia salmonis
CC comprising administering either intraperitoneally, by immersion or
CC orally, an immunogenic amount of principal antigen, the OspA (outer
CC surface lipoprotein), its variants, non-lipidated form or antigenic
CC peptides derived or synthesized with or without an adjuvant. The new
CC method is used to provide an outer surface lipoprotein (OspA) of
CC bacterial strain Piscirickettsia salmonis as a vaccine to induce immunity
CC in fin-fish against Rickettsial septicaemia and other related
CC Rickettsial diseases caused by either a virus, bacteria or parasite.
CC This sequence represents an oligonucleotide used to construct an
CC Escherichia coli codon optimised outer surface lipoprotein OspA (17e2)
CC used in the creation of the vaccine described in the invention.
XX
SQ Sequence 110 BP; 13 A; 28 C; 45 G; 24 T; 0 other;

Query Match 22.4%; Score 108.4; DB 24; Length 110;
Best Local Similarity 99.1%; Pred. No. 5,5e-23;
Matches 109; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 53 TGGGTTGGCCCGCCAGAACTTCAGCCGCCAGGAAGTTGGCGCGCCACCGGTGCGGTGTGG 112
|||||
Db 1 TGGGTTGGCCCGCCAGAACTTCAGCCGCCAGGAAGTTGGCGCGCCACCGGTGCGGTGTGG 60

QY 113 GCGGTTGGCCCGCCAGCTGTTCCGTTAAAGGCTCTGGTCTGTCTGTCGATG 162
|||||
Db 61 GCGGTTGGCCCGCCAGCTGTTCCGTTAAAGGCTCTGGTCTGTCTGTCGATG 110

RESULT 12
AAF86255/c
ID AAF86255 standard; DNA; 102 BP.
XX
AC AAF86255;
XX
XX 11-JUL-2001 (first entry)
XX
XX PCR primer #5 used in cloning an optimisation of OspA gene.
XX
KW Poikilothermic fish; Piscirickettsia salmonis; rickettsial pathogen;
KW vaccine; OspA; salmonid rickettsial septicaemia; rickettsial disease;
KW SRS; 17E2; fusion construct; PCR primer; ss.
XX
OS Piscirickettsia salmonis.
XX
PN CA2281913-Al.
XX
PD 17-MAR-2001.
XX
XX 17-SEP-1999; 99CA-2281913.
PF 17-SEP-1999; 99CA-2281913.
PR 17-SEP-1999; 99CA-2281913.
XX
XX (KAYW/) KAY W W.
PA (BURL/) BURIAN J.
PA (KUZY/) KUZIK M A.
XX
XX Kay WW, Burian J, Kuzyk MA;
XX
XX WPI; 2001-316844/34.
XX
XX Method for protecting poikilothermic fish against salmonid rickettsial
PT septicaemia and other rickettsial diseases comprises administering a
PT vaccine containing the OspA protein of Piscirickettsia salmonis -
XX
XX Example 3; Fig 4B; 35pp; English.
XX
XX This invention relates to a method for the protection against infection
CC of a poikilothermic fish by the bacterial pathogen, Piscirickettsia
CC salmonis. The method comprises administering an immunogenic amount of a
CC P. salmonis specific antigen termed OspA, or an immunogenic fragment of
CC OspA in the form of a vaccine. The method is used for protecting animals,
CC particularly poikilothermic fish, against the bacterial pathogen
CC P. salmonis. The method is also useful for protecting against salmonid
CC rickettsial septicaemia (SRS) and other rickettsial diseases. The present
CC sequence represents a PCR primer used in the cloning and optimisation of
CC the P. salmonis OspA gene. The OspA gene is used in the method of the
CC invention.
XX
SQ Sequence 102 BP; 14 A; 23 C; 30 G; 35 T; 0 other;

Query Match 21.1%; Score 102; DB 22; Length 102;
Best Local Similarity 100.0%; Pred. No. 4.6e-21;
Matches 102; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 334 ACCTACCGCGTTACACAAACAGGACCGGTCAGCAGTACTGCGCGGAATTCAGCAG 393
|||||
Db 102 ACCTACCGCGTTACACAAACAGGACCGGTCAGCAGTACTGCGCGGAATTCAGCAG 43

QY 394 AAAGCCATCATCGAGGTCAGAAACAGGAAGTAATCTACGGCACCC 435
|||||
Db 42 AAAGCCATCATCGAGGTCAGAAACAGGAAGTAATCTACGGCACCC 1

RESULT 13
ABK52410/c
ID ABK52410 standard; DNA; 102 BP.

PN CA2339327-A1.
XX
PD 15-MAR-2002.
XX
PF 19-MAR-2001; 2001CA-2339327.
XX
PR 15-SEP-2000; 2000US-0677374.
XX
PA (THOR/) THORNTON J C.
PA (KAYW/) KAY W W.
PA (BURI/) BURIAN J.
PA (KUZY/) KUZYSK M A.
XX
PI Thornton JC, Kay WW, Burian J, Kuzyk MA;
XX WPI; 2002-45221/49.
XX
XX
PT Inducing immunity in fin fish to Rickettsial septicaemia, comprises
PT administration of an outer surface lipoprotein (OspA) of a bacterial
PT strain, as a vaccine -
XX
XX
PS Example 3; Fig 4; 55pp; English.
XX
XX The invention describes a method of protecting a poikilothermic fish
CC against infection by the bacterial pathogen *Piscirickettsia salmonis*
CC comprising administering either intraperitoneally, by immersion or
CC orally, an immunogenic amount of principal antigen, the OspA (outer
CC surface lipoprotein), its variants, non-lipidated form or antigenic
CC peptides derived or synthesized with or without an adjuvant. The new
CC method is used to provide an outer surface lipoprotein (OspA) of
CC bacterial strain *Piscirickettsia salmonis* as a vaccine to induce immunity
CC in fin-fish against Rickettsial septicaemia and other related
CC Rickettsial diseases caused by either a virus, bacteria or parasite.
CC This sequence represents an oligonucleotide used to construct an
CC *Escherichia coli* codon optimised outer surface lipoprotein OspA (17e2)
CC used in the creation of the vaccine described in the invention.
XX
SQ Sequence 94 BP; 18 A; 20 C; 35 G; 21 T; 0 other;

Query Match 19.1%; Score 92.4; DB 24; Length 94;
Best Local Similarity 98.9%; Pred. No. 3.6e-18;
Matches 93; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 139 AAAGGCTCTGGTGTGTCGATGCCATCGCGGTCTGGGGTCTGATTGGC 198
DB 1 AAAGGCTCTGGTGTGTCGATGCCATCGCGGTCTGGGGTCTGATTGGC 60
QY 199 TCTAAATCGGTCAGACATGGACGACGAGGATA 232
DB 61 TCTAAATCGGTCAGACATGGACGACGAGGATA 94

Search completed: February 22, 2003, 14:44:57
Job time : 141.862 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 22, 2003, 14:26:45 ; Search time 957.936 Seconds
(without alignments)
8165.910 Million cell updates/sec

Title: US-09-677-374-3

Perfect score: 483

Sequence: 1 atgcgtgttgctgcagg.....agtgattagaccgaaaaa 483

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST: *
1: em_estba: *
2: em_esthum: *
3: em_estin: *
4: em_estinu: *
5: em_estov: *
6: em_estpl: *
7: em_estro: *
8: em_hc: *
9: gb_est1: *
10: gb_est2: *
11: gb_hc: *
12: gb_est3: *
13: gb_est4: *
14: gb_est5: *
15: em_estfun: *
16: em_estom: *
17: gb_gss: *
18: em_gss_hum: *
19: em_gss_inv: *
20: em_gss_pin: *
21: em_gss_vrt: *
22: em_gss_fun: *
23: em_gss_mam: *
24: em_gss_mus: *
25: em_gss_other: *
26: em_gss_pro: *
27: em_gss_rod: *

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|-------------|
| 1 | 42 | 8.7 | 238 | 12 | BG240789 |
| 2 | 42 | 8.7 | 402 | 12 | BG356895 |
| 3 | 42 | 8.7 | 419 | 10 | BE355894 |
| 4 | 42 | 8.7 | 428 | 12 | BF176742 |
| 5 | 42 | 8.7 | 446 | 12 | BG052290 |
| 6 | 42 | 8.7 | 500 | 12 | BF586874 |

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

| | | | | | |
|------|------|-----|------|----|----------|
| 7 | 42 | 8.7 | 507 | 12 | BG053591 |
| 8 | 42 | 8.7 | 531 | 12 | BG053014 |
| 9 | 42 | 8.7 | 537 | 12 | BG713844 |
| 10 | 42 | 8.7 | 561 | 12 | BF587805 |
| 11 | 42 | 8.7 | 577 | 12 | BG673858 |
| 12 | 42 | 8.7 | 582 | 10 | BE356763 |
| 13 | 42 | 8.7 | 591 | 10 | BE356800 |
| 14 | 42 | 8.7 | 597 | 12 | BG102589 |
| 15 | 42 | 8.7 | 598 | 10 | AW672446 |
| 16 | 42 | 8.7 | 600 | 10 | BE360868 |
| 17 | 42 | 8.7 | 620 | 12 | BF481524 |
| 18 | 42 | 8.7 | 649 | 10 | BE355895 |
| 19 | 40.4 | 8.4 | 548 | 13 | BM328366 |
| 20 | 40.2 | 8.3 | 508 | 14 | BQ242846 |
| C 21 | 40 | 8.3 | 513 | 17 | AV430609 |
| C 22 | 40 | 8.3 | 768 | 17 | AQ690676 |
| C 23 | 38 | 7.9 | 910 | 17 | CNS0060N |
| C 24 | 38 | 7.9 | 1101 | 17 | CNS0100Y |
| C 25 | 37.8 | 7.8 | 595 | 13 | BI607335 |
| C 26 | 37.8 | 7.8 | 826 | 13 | BI522679 |
| C 27 | 37.8 | 7.8 | 925 | 17 | CNS0091P |
| C 28 | 37.6 | 7.8 | 1344 | 12 | BE964680 |
| C 29 | 37.4 | 7.7 | 350 | 12 | BG052289 |
| C 30 | 37.4 | 7.7 | 1101 | 17 | CNS0178Y |
| C 31 | 37.2 | 7.7 | 399 | 14 | BQ753608 |
| C 32 | 37.2 | 7.7 | 444 | 14 | BQ753422 |
| C 33 | 37.2 | 7.7 | 514 | 12 | BG320949 |
| C 34 | 37 | 7.7 | 908 | 12 | BF253465 |
| C 35 | 36.8 | 7.6 | 446 | 9 | AU182544 |
| C 36 | 36.8 | 7.6 | 1101 | 9 | AL513871 |
| C 37 | 36.4 | 7.5 | 604 | 14 | BM936458 |
| C 38 | 36.4 | 7.5 | 1043 | 12 | BG104511 |
| C 39 | 36.2 | 7.5 | 595 | 17 | BH478914 |
| C 40 | 36.2 | 7.5 | 647 | 13 | BJ003370 |
| C 41 | 36.2 | 7.5 | 663 | 13 | BJ003303 |
| C 42 | 36.2 | 7.5 | 789 | 9 | AU125061 |
| C 43 | 36 | 7.5 | 1058 | 12 | BG295311 |
| C 44 | 35.8 | 7.4 | 682 | 10 | BE593319 |
| C 45 | 35.6 | 7.4 | 432 | 14 | BM821964 |

ALIGNMENTS

| | | | | |
|------------|--|--------|--------|----------------------------|
| RESULT 1 | BG240789 | 238 bp | linear | EST 15-FEB-2001 |
| LOCUS | OV1_38_C06.g1_A002 Ovary 1 (OV1) | 238 bp | mRNA | Sorghum bicolor cdNA, mRNA |
| DEFINITION | Sequence. | | | |
| ACCESSION | BG240789 | | | |
| VERSION | BG240789.1 | | | |
| KEYWORDS | EST. | | | |
| SOURCE | Sorghum. | | | |
| ORGANISM | Sorghum bicolor | | | |
| REFERENCE | Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoideae; Andropogoneae; Sorghum. | | | |
| AUTHORS | 1 (bases 1 to 238) | | | |
| TITLE | Cordonnier-Pratt, M.-M., Gingle, A., Marsala, C., Sudman, M. and Pratt, J.L.H. | | | |
| JOURNAL | An EST database from Sorghum: ovaries of varying immature stages | | | |
| COMMENT | Unpublished (2000) | | | |
| | Contact: Cordonnier-Pratt MM | | | |
| | Laboratory for Genomics and Bioinformatics | | | |
| | The University of Georgia, Department of Plant Biology | | | |
| | Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA | | | |
| | Tel: 706 542 1860 | | | |
| | Fax: 706 583 0210 | | | |
| | Email: mmpratt@uga.edu | | | |
| | Sequences have been trimmed to exclude PolyA, vector and regions below Phred quality 16. The threshold for highest quality sequence is 20. | | | |
| | Seq primer: PolyTmix | | | |


```

Db 138 CGACGGCCAGCACTATGAGAGTTTGTAAAGTTATGATGCCCAAGTGAGGAGCGGT 197
QY 180 TC 181
Db 198 CC 199

RESULT 4
BF176742
LOCUS
DEFINITION
EM1_4_F01_g1_A002 Embryo 1 (EM1) Sorghum bicolor cDNA, mRNA
VERSION
BF176742
KEYWORDS
SOURCE
ORGANISM
Sorghum bicolor
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
Clade; Panicoideae; Andropogoneae; Sorghum.
1 (bases 1 to 428)
Reid,S.P., Cordonnier-Pratt,M.-M., Gingle,A. and Pratt,L.H.
An EST database from Sorghum: developing embryos
Unpublished (2000)
Contact: Cordonnier-Pratt MM
Laboratory for Genomics and Bioinformatics
The University of Georgia, Department of Plant Biology
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
Tel: 706 542 1860
Fax: 706 583 0210
Email: mmpratt@uga.edu
Sequences have been trimmed to exclude PolyA, vector and regions
below Phred quality 16. The threshold for highest quality sequence
is 20.
Seq primer: PolyTMix
High quality sequence start: 32
High quality sequence stop: 427
POLYA-No.
FEATURES
source
1..428
/organism="Sorghum bicolor"
/db_xref="taxon:4558"
/clone_lib="Embryo 1 (EM1)"
/note="Organ: Embryos germinated for 24 hr; Vector:
pBluescript II from Lambda Zap II; Site_1: XhoI; Site_2:
EcoRI; The library was made from poly-A RNA in the cloning
vector lambda Zap II. Clones to be sequenced were
prepared by mass excision."
BASE COUNT 103 a 80 c 119 g 126 t
ORIGIN

Query Match 8.7%; Score 42; DB 12; Length 428;
Best Local Similarity 59.0%; Pred. No. 0.036;
Matches 72; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

QY 60 CGCCGAGAACTTCAGCGCCAGGAAGTTGGCGGCGCCACCGGTTCGTGGCGGTGT 119
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 58 CGCGGAGAACTGACCCAGCAGGAGTTCGACGAGATGATCCGTGAGGCTGACGTGATGG 117
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 120 TGGCGGCACGTTCGGTAAAGCTCTGCTGCTGATGCCATCGCGGTGCGGT 179
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 118 CGACGGCCAGCACTATGAGAGTTTGTAAAGTTATGATGCCCAAGTGAGGAGCGGT 177
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 180 TC 181
Db 178 CC 179

RESULT 5
BG052290
LOCUS
DEFINITION
RH122_12.G04.g1_A003 Rhizome2 (RH122) Sorghum propinquum cDNA, mRNA
sequence.
BG052290
LOCUS
DEFINITION
EM1_31_E06_g1_A003 Floral-Induced Meristem 1 (FMI) Sorghum
propinquum cDNA, mRNA sequence.
BF586874
VERSION
BF586874
KEYWORDS
SOURCE
ORGANISM
Sorghum propinquum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
Clade; Panicoideae; Andropogoneae; Sorghum.
1 (bases 1 to 500)
Cordonnier-Pratt,M.-M., Gingle,A., Sudman,M., Marsala,C. and Pratt
,L.H.
An EST database from Sorghum: floral-induced meristems
Unpublished (2000)
Contact: Cordonnier-Pratt MM
Laboratory for Genomics and Bioinformatics

```

```

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Sorghum propinquum.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
Clade; Panicoideae; Andropogoneae; Sorghum.
1 (bases 1 to 446)
Cordonnier-Pratt,M.-M., Gingle,A., Paterson,A., Sudman,M. and Pratt
,L.H.
An EST database from Sorghum: Sorghum propinquum rhizomes
Unpublished (2000)
Contact: Cordonnier-Pratt MM
Laboratory for Genomics and Bioinformatics
The University of Georgia, Department of Plant Biology
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
Tel: 706 542 1860
Fax: 706 583 0210
Email: mmpratt@uga.edu
Seq primer: PolyTMix
High quality sequence start: 11
High quality sequence stop: 440
POLYA-No.
FEATURES
source
1..446
/organism="Sorghum propinquum"
/db_xref="taxon:132711"
/clone_lib="Rhizome2 (RH122)"
/note="Organ: Rhizomes; Vector: pBluescript II from Lambda
Zap II; Site_1: XhoI; Site_2: EcoRI; The library was made
from poly-A RNA in the cloning vector lambda Zap II.
Clones to be sequenced were prepared by mass excision."
BASE COUNT 101 a 81 c 124 g 140 t
ORIGIN

Query Match 8.7%; Score 42; DB 12; Length 446;
Best Local Similarity 59.0%; Pred. No. 0.037;
Matches 72; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

QY 60 CGCCGAGAACTTCAGCGCCAGGAAGTTGGCGGCGCCACCGGTTCGTGGCGGTGT 119
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 51 CGGCGAAGCTGACCCAGCAGGAGTTCGACGAGATGATCCGTGAGGCTGACGTGATGG 110
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 120 TGGCGGCACGTTCGGTAAAGCTCTGCTGCTGATGCCATCGCGGTGCGGT 179
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 111 CGACGGCCAGCACTATGAGAGTTTGTAAAGTTATGATGCCCAAGTGAGGAGCGGT 170
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 180 TC 181
Db 171 CC 172

RESULT 6
BF586874
LOCUS
DEFINITION
FMI_31_E06_g1_A003 Floral-Induced Meristem 1 (FMI) Sorghum
propinquum cDNA, mRNA sequence.
BF586874
VERSION
BF586874
KEYWORDS
SOURCE
ORGANISM
Sorghum propinquum.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
Clade; Panicoideae; Andropogoneae; Sorghum.
1 (bases 1 to 500)
Cordonnier-Pratt,M.-M., Gingle,A., Sudman,M., Marsala,C. and Pratt
,L.H.
An EST database from Sorghum: floral-induced meristems
Unpublished (2000)
Contact: Cordonnier-Pratt MM
Laboratory for Genomics and Bioinformatics

```

The University of Georgia, Department of Plant Biology
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
Tel: 706 542 1860
Fax: 706 583 0210
Email: mmpratt@uga.edu
Sequences have been trimmed to exclude PolyA, vector and regions
below Phred quality 16. The threshold for highest quality sequence
is 20.

Seq primer: PolyTMix
High quality sequence start: 8
High quality sequence stop: 495
POLYA-No.

FEATURES source Location/Qualifiers
1. .500

/organism="Sorghum propinquum"

/db_xref="taxon:132711"

/clone_lib="Floral-Induced Meristem 1 (FM1)"

/note="Organ: Floral-induced meristems; Vector:

pBluescript II from Lambda Zap II; Site_1: XhoI; Site_2:
EcoRI; mature plants were placed in a growth chamber for
15 days with 16 hr darkness and 8 hr light (flowering is
induced by short-day conditions); 16 days after being
returned to the greenhouse under natural long days during
late April/early May, meristems were harvested. The
library was made from poly-A RNA in the cloning vector
lambda ZAP II. Clones to be sequenced were prepared by
mass excision."

BASE COUNT 111 a 94 c 141 g 154 t

ORIGIN

Query Match 8.7%; Score 42; DB 12; Length 500;

Best Local Similarity 59.0%; Pred. No. 0.039;

Matches 72; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

QY 60 GCCCGAGAACTTCAGCGCCAGGAGTTGGCGGCCACCGGTGTGTGGCGGGTGT 119

DB 89 CGCGGAGAGCTGACCGAGGAGGTGCGACGAGATGATCCGTGAGGCTGACGTGATGG 148

QY 120 TCCCGGCCAGCTGTTCCGTTAAAGGCTCTGGTGTGTGATGGCCATCGCGCGGTGCGGT 179

DB 149 CGACGCCAGATCAACTATGAAGATTGTTAAAGTTATGATGGCCAAAGTGAGGAGCGGT 208

QY 180 TC 181

DB 209 CC 210

RESULT 7

BG053591

LOCUS RH122_11_C03_g1_A003 Rhizome2 (RH122) Sorghum propinquum cDNA, mRNA
DEFINITION

sequence.

ACCESSION BG053591

VERSION BG053591.1 GI:12509440

KEYWORDS EST.

SOURCE Sorghum propinquum.

ORGANISM Sorghum propinquum

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
clade; Panicoideae; Andropogoneae; Sorghum.

REFERENCE 1 (bases 1 to 507)

AUTHORS Cordonnier-Pratt,M.-M., Gingle,A., Paterson,A., Sudman,M. and Pratt
L.H.

TITLE An EST database from Sorghum: Sorghum propinquum rhizomes

JOURNAL Unpublished (2000)

COMMENT Contact: Cordonnier-Pratt MM

Laboratory for Genomics and Bioinformatics

The University of Georgia, Department of Plant Biology

Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA

Tel: 706 542 1860

Fax: 706 583 0210

Email: mmpratt@uga.edu

Seq primer: 17

High quality sequence start: 95
High quality sequence stop: 507
POLYA=Yes.

FEATURES source Location/Qualifiers
1. .507

/organism="Sorghum propinquum"

/db_xref="taxon:132711"

/clone_lib="Rhizome2 (RH122)"

/note="Organ: Rhizomes; Vector: pBluescript II from Lambda
Zap II; Site_1: XhoI; Site_2: EcoRI; The library was made
from poly-A RNA in the cloning vector lambda ZAP II.
Clones to be sequenced were prepared by mass excision."

BASE COUNT 109 a 92 c 142 g 164 t

ORIGIN

Query Match 8.7%; Score 42; DB 12; Length 507;

Best Local Similarity 59.0%; Pred. No. 0.04;

Matches 72; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

QY 60 GCCCGAGAACTTCAGCGCCAGGAGTTGGCGGCCACCGGTGTGTGGCGGGTGT 119

DB 54 CGCGGAGAGCTGACCGAGGAGGTGCGACGAGATGATCCGTGAGGCTGACGTGATGG 113

QY 120 TCCCGGCCAGCTGTTCCGTTAAAGGCTCTGGTGTGTGATGGCCATCGCGCGGTGCGGT 179

DB 114 CGACGCCAGATCAACTATGAAGATTGTTAAAGTTATGATGGCCAAAGTGAGGAGCGGT 173

QY 180 TC 181

DB 174 CC 175

RESULT 8

BG053014

LOCUS RH122_16_D04_g1_A003 Rhizome2 (RH122) Sorghum propinquum cDNA, mRNA

DEFINITION

sequence.

ACCESSION BG053014

VERSION BG053014.1 GI:12508270

KEYWORDS EST.

SOURCE Sorghum propinquum.

ORGANISM Sorghum propinquum

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC

clade; Panicoideae; Andropogoneae; Sorghum.

REFERENCE 1 (bases 1 to 531)

AUTHORS Cordonnier-Pratt,M.-M., Gingle,A., Paterson,A., Sudman,M. and Pratt

L.H.

TITLE An EST database from Sorghum: Sorghum propinquum rhizomes

JOURNAL Unpublished (2000)

COMMENT Contact: Cordonnier-Pratt MM

Laboratory for Genomics and Bioinformatics

The University of Georgia, Department of Plant Biology

Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA

Tel: 706 542 1860

Fax: 706 583 0210

Email: mmpratt@uga.edu

Seq primer: PolyTMix

High quality sequence start: 42

High quality sequence stop: 527

POLYA-No.

FEATURES source Location/Qualifiers

1. .531

/organism="Sorghum propinquum"

/db_xref="taxon:132711"

/clone_lib="Rhizome2 (RH122)"

/note="Organ: Rhizomes; Vector: pBluescript II from Lambda

Zap II; Site_1: XhoI; Site_2: EcoRI; The library was made

from poly-A RNA in the cloning vector lambda ZAP II.

Clones to be sequenced were prepared by mass excision."

BASE COUNT 122 a 101 c 152 g 154 t

ORIGIN

```

Query Match      8.7%; Score 42; DB 12; Length 531;
Best Local Similarity 59.0%; Pred. No. 0.04;
Matches 72; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

QY 60 CGCCAGAACTTCAGCGCCAGGAAGTTGGCGCCGACCGGTGCGTGTGTGGCGGTGT 119
DB 121 CGGCGAGAACTGACCCAGCAGGAGTTCGAGGAGATATCGGTGAGGCTGACGTCGATGG 180
QY 120 TGGCGGCCAGCTGTTCGGTAAAGCTCTGCTGTGTGTCGATGCCATCGCGGTGCGGT 179
DB 181 CGACGGCCAGATCAACTATGAGAGTTTGTAAAGTTATGATGCCCAAGTGAGGAGCGGT 240
QY 180 TC 181
DB 241 CC 242

RESULT 9
LOCUS BG713844 537 bp mRNA linear EST 08-MAY-2001
DEFINITION EMI_20_C08_g2_A002 Embryo 1 (EM1) Sorghum bicolor cDNA, mRNA
VERSION BG713844
KEYWORDS BG713844.1 GI:14007794
SOURCE EST.
ORGANISM Sorghum bicolor
REFERENCE Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
AUTHORS Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
TITLE clade; Panicoideae; Andropogoneae; Sorghum.
JOURNAL 1 (bases 1 to 537)
COMMENT Reid,S.P., Cordonnier-Pratt,M.-M., Gingle,A. and Pratt,L.H.
An EST database from Sorghum: developing embryos
Contact: Cordonnier-Pratt MM
Laboratory for Genomics and Bioinformatics
The University of Georgia, Department of Plant Biology
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
Tel: 706 542 1860
Fax: 706 583 0210
Email: mmpratt@uga.edu
Sequences have been trimmed to exclude PolyA, vector and regions
below Phred quality 16. The threshold for highest quality sequence
is 20.
Seq primer: PolyTMix
High quality sequence start: 62
High quality sequence stop: 536
POLYA-No.
Location/Qualifiers
1. .537
/organism="Sorghum bicolor"
/db_xref="taxon:4558"
/clone.lib="Embryo 1 (EM1)"
/notes="Organ: Embryos germinated for 24 hr; Vector:
pBluescript II from Lambda Zap II; Site_1: XhoI; Site_2:
EcoRI; The library was made from poly-A RNA in the cloning
vector lambda Zap II. Clones to be sequenced were
prepared by mass excision."
BASE COUNT 118 a 98 c 152 g 169 t
ORIGIN
Query Match      8.7%; Score 42; DB 12; Length 537;
Best Local Similarity 59.0%; Pred. No. 0.041;
Matches 72; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

QY 60 CGCCAGAACTTCAGCGCCAGGAAGTTGGCGCCGACCGGTGCGTGTGTGGCGGTGT 119
DB 92 CGGCGAGAACTGACCCAGCAGGAGTTCGAGGAGATATCGGTGAGGCTGACGTCGATGG 151
QY 120 TGGCGGCCAGCTGTTCGGTAAAGCTCTGCTGTGTGTCGATGCCATCGCGGTGCGGT 179
DB 152 CGACGGCCAGATCAACTATGAGAGTTTGTAAAGTTATGATGCCCAAGTGAGGAGCGGT 211

Query Match      8.7%; Score 42; DB 12; Length 561;
Best Local Similarity 59.0%; Pred. No. 0.042;
Matches 72; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

QY 60 CGCCAGAACTTCAGCGCCAGGAAGTTGGCGCCGACCGGTGCGTGTGTGGCGGTGT 119
DB 202 CGGCGAGAACTGACCCAGCAGGAGTTCGAGGAGATATCGGTGAGGCTGACGTCGATGG 261
QY 120 TGGCGGCCAGCTGTTCGGTAAAGCTCTGCTGTGTGTCGATGCCATCGCGGTGCGGT 179
DB 262 CGACGGCCAGATCAACTATGAGAGTTTGTAAAGTTATGATGCCCAAGTGAGGAGCGGT 321
QY 180 TC 181
DB 322 CC 323

RESULT 11

```

BG673858
LOCUS EML20.C08.g1_A002 Embryo 1 (EML1) Sorghum bicolor cDNA, mRNA EST 01-MAY-2001
DEFINITION sequence.
ACCESSION BG673858
VERSION BG673858.1 GI:13919511
KEYWORDS EST.
SOURCE sorghum.
ORGANISM Sorghum bicolor
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
clade; Panicoideae; Andropogoneae; Sorghum.
1 (bases 1 to 577)
Reid,S.P., Cordonnier-Pratt,M.-M., Gingle,A. and Pratt,L.H.
TITLE An EST database from Sorghum: developing embryos
JOURNAL Unpublished (2000)
COMMENT Contact: Cordonnier-Pratt MM
Laboratory for Genomics and Bioinformatics
The University of Georgia, Department of Plant Biology
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
Tel: 706 542 1860
Fax: 706 583 0210
Email: mmpratt@uga.edu
Sequences have been trimmed to exclude polyA, vector and regions
below Phred quality 16. The threshold for highest quality sequence
is 20.
Seq primer: PolyTMix
High quality sequence start: 57
High quality sequence stop: 572
POLYA-No.
FEATURES
source
1..577
/organism="Sorghum bicolor"
/db_xref="taxon:4558"
/clone_lib="Embryo 1 (EML1)"
/note="Organ: Embryos germinated for 24 hr; Vector:
pBluescript II from lambda zap II; Site_1: XhoI; Site_2:
EcoRI; The library was made from poly-A RNA in the cloning
vector lambda ZAP II. Clones to be sequenced were
prepared by mass excision."
BASE COUNT 129 a 106 c 166 g 176 t
ORIGIN
Query Match 8.7%; Score 42; DB 12; Length 577;
Best Local Similarity 59.0%; Pred. NO. 0.042;
Matches 72; Conservative 0; Mismatches 50; Indels 0; Gaps 0;
QY 60 CGCCGAGAACTTCAGCCGCCAGGAGTTGGCGGCCACCGCTGGGTGTGGCGGTGT 119
Db 121 CGCGGAGAGCTGACCGGAGGAGGTGCGACGAGATGATCCGTGAGGCTGCGATGG 180
QY 120 TGCCGCCAGCTGTTCGGTAAAGGCTCTGGTGTGTGTCGATGGCCATCGCGGTGCGGT 179
Db 181 CGACGCCAGATCACTATGAGAGCTTCTTAAGGTTATGATGATGCCAAGTGAGGACGCT 240
QY 180 TC 181
Db 241 CC 242
RESULT 12
BE356763
LOCUS DGL12.B12.g1_A002 Dark Grown 1 (DGL1) Sorghum bicolor cDNA, mRNA EST 20-JUL-2000
DEFINITION sequence.
ACCESSION BE356763
VERSION BE356763.1 GI:9298320
KEYWORDS EST.
SOURCE sorghum.
ORGANISM Sorghum bicolor
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
clade; Panicoideae; Andropogoneae; Sorghum.

1 (bases 1 to 582)
Cordonnier-Pratt,M.-M., Gingle,A., Marsala,C., Sudman,M. and Pratt
L.H.
TITLE An EST database from Sorghum: dark-grown seedlings
JOURNAL Unpublished (2000)
COMMENT Contact: Cordonnier-Pratt MM
Laboratory for Genomics and Bioinformatics
The University of Georgia, Department of Plant Biology
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
Tel: 706 542 1860
Fax: 706 583 0210
Email: mmpratt@uga.edu
Sequences have been trimmed to exclude polyA, vector and regions
below Phred quality 16. The threshold for highest quality sequence
is 20.
Seq primer: PolyTMix
High quality sequence start: 6
High quality sequence stop: 553
POLYA-No.
FEATURES
source
1..582
/organism="Sorghum bicolor"
/db_xref="taxon:4558"
/clone_lib="Dark Grown 1 (DGL1)"
/note="Organ: 5-day-old dark-grown seedlings; Vector:
lambda Zap; Site_1: XhoI; Site_2: EcoRI; The library was
made from poly-A RNA in the cloning vector lambda ZAP II.
Clones to be sequenced were prepared by mass excision."
BASE COUNT 145 a 116 c 166 g 155 t
ORIGIN
Query Match 8.7%; Score 42; DB 10; Length 582;
Best Local Similarity 59.0%; Pred. NO. 0.042;
Matches 72; Conservative 0; Mismatches 50; Indels 0; Gaps 0;
QY 60 CCCCAGAACTTCAGCCGCCAGGAGTTGGCGGCCACCGCTGGGTGTGGCGGTGT 119
Db 213 CGCGGAGAGCTGACCGGAGGAGGTGCGACGAGATGATCCGTGAGGCTGCGATGG 272
QY 120 TGCCGCCAGCTGTTCGGTAAAGGCTCTGGTGTGTGTCGATGGCCATCGCGGTGCGGT 179
Db 273 CGACGCCAGATCACTATGAGAGCTTCTTAAGGTTATGATGATGCCAAGTGAGGACGCT 332
QY 180 TC 181
Db 333 CC 334
RESULT 13
BE356800
LOCUS DGL12.E10.g1_A002 Dark Grown 1 (DGL1) Sorghum bicolor cDNA, mRNA EST 20-JUL-2000
DEFINITION sequence.
ACCESSION BE356800
VERSION BE356800.1 GI:9298357
KEYWORDS EST.
SOURCE sorghum.
ORGANISM Sorghum bicolor
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
clade; Panicoideae; Andropogoneae; Sorghum.
1 (bases 1 to 591)
Cordonnier-Pratt,M.-M., Gingle,A., Marsala,C., Sudman,M. and Pratt
L.H.
TITLE An EST database from Sorghum: dark-grown seedlings
JOURNAL Unpublished (2000)
COMMENT Contact: Cordonnier-Pratt MM
Laboratory for Genomics and Bioinformatics
The University of Georgia, Department of Plant Biology
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
Tel: 706 542 1860
Fax: 706 583 0210
Email: mmpratt@uga.edu

Sequences have been trimmed to exclude polyA, vector and regions below Phred quality 16. The threshold for highest quality sequence is 20.

Seq primer: PolyTMix
High quality sequence start: 10
High quality sequence stop: 557
POLYA-No.

FEATURES

source
1. .591
Location/Qualifiers
/organism="Sorghum bicolor"
/db_xref="taxon:4558"
/clone_lib="Dark Grown 1 (DGI)"
/note="Organ: 5-day-old dark-grown seedlings; Vector:
Lambda Zap; Site_1: XhoI; Site_2: EcoRI; The library was
made from poly-A RNA in the cloning vector lambda ZAP II.
Clones to be sequenced were prepared by mass excision."

BASE COUNT 146 a 114 c 168 g 163 t
ORIGIN
Query Match 8.7%; Score 42; DB 10; Length 591;
Best Local Similarity 59.0%; Pred. No. 0.043;
Matches 72; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

QY 60 CGCCAGAACTTCAGCGCCAGGAAGTTGGCGGCCACCGTGGTGTGGCGGTGT 119
DB 210 CGCGAGAGCTGACCGACGAGGAGTTCGACGAGATGATCGTGAGCTGACGTGATGG 269
QY 120 TGGCGGCCAGCTGTTCGGTAAAGCTCTGTCGTGTCGATGCCATCGCGTGGCGGT 179
DB 270 CGACGCCAGATCAACTATGAAGAGTTTGTAAAGTTATGATGCCCAAGTGAAGGCGGT 329
QY 180 TC 181
DB 330 CC 331

RESULT 14
LOCUS BG102589 597 bp mRNA linear EST 30-JAN-2001
DEFINITION RHIZ2_34_H05.g1_A003 Rhizome2 (RHIZ2) Sorghum propinquum cDNA, mRNA
sequence.
ACCESSION BG102589
VERSION BG102589.1 GI:12617422
KEYWORDS EST.
SOURCE Sorghum propinquum.
ORGANISM Sorghum propinquum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
clade; Panicoideae; Andropogoneae; Sorghum.

REFERENCE 1 (bases 1 to 597)
AUTHORS Cordonnier-Pratt, M.-M., Gingle, A., Paterson, A., Sudman, M. and Pratt, L.H.
TITLE An EST database from Sorghum: Sorghum propinquum rhizomes
JOURNAL Unpublished (2000)
COMMENT Contact: Cordonnier-Pratt MM
Laboratory for Genomics and Bioinformatics
The University of Georgia, Department of Plant Biology
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
Tel: 706 542 1860
Fax: 706 583 0210
Email: mmpratt@uga.edu

Seq primer: PolyTMix
High quality sequence start: 59
High quality sequence stop: 593
POLYA-No.

FEATURES

source
1. .597
Location/Qualifiers
/organism="Sorghum propinquum"
/db_xref="taxon:132711"
/clone_lib="Rhizome2 (RHIZ2)"
/note="Organ: Rhizomes; Vector: pBluescript II from Lambda
Zap II; Site_1: XhoI; Site_2: EcoRI; The library was made
from poly-A RNA in the cloning vector lambda ZAP II.

BASE COUNT 149 a 130 c 170 g 148 t
ORIGIN
Clones to be sequenced were prepared by mass excision."

Query Match 8.7%; Score 42; DB 12; Length 597;
Best Local Similarity 59.0%; Pred. No. 0.043;
Matches 72; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

QY 60 CGCCAGAACTTCAGCGCCAGGAAGTTGGCGGCCACCGTGGTGTGGCGGTGT 119
DB 267 CGCGAGAGCTGACCGACGAGGAGTTCGACGAGATGATCGTGAGCTGACGTGATGG 326
QY 120 TGGCGGCCAGCTGTTCGGTAAAGCTCTGTCGTGTCGATGCCATCGCGTGGCGGT 179
DB 327 CGACGCCAGATCAACTATGAAGAGTTTGTAAAGTTATGATGCCCAAGTGAAGGCGGT 386
QY 180 TC 181
DB 387 CC 388

RESULT 15
LOCUS AW672446 598 bp mRNA linear EST 19-JUL-2000
DEFINITION LG1_359_D09.g1_A002 Light Grown 1 (LG1) Sorghum bicolor cDNA, mRNA
sequence.

ACCESSION AW672446
VERSION AW672446.1 GI:7536365
KEYWORDS EST.
SOURCE sorghum.
ORGANISM Sorghum bicolor

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
clade; Panicoideae; Andropogoneae; Sorghum.

REFERENCE 1 (bases 1 to 598)
AUTHORS Cordonnier-Pratt, M.-M., Gingle, A., Marsala, C. and Pratt, L.H.
TITLE An EST database from Sorghum: light-grown seedlings
JOURNAL Unpublished (2000)
COMMENT Contact: Cordonnier-Pratt MM
Laboratory for Genomics and Bioinformatics
The University of Georgia, Department of Plant Biology
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
Tel: 706 542 1860
Fax: 706 583 0210
Email: mmpratt@uga.edu

Sequences have been trimmed to exclude polyA, vector and regions below Phred quality 16. The threshold for highest quality sequence is 20.

Seq primer: T7

High quality sequence start: 71
High quality sequence stop: 598
POLYA=Yes.

FEATURES

source
1. .598
Location/Qualifiers
/organism="Sorghum bicolor"
/db_xref="taxon:4558"
/clone_lib="Light Grown 1 (LG1)"
/note="Organ: 10- to 14-day-old light-grown (greenhouse)
seedlings; Vector: Lambda Zap; Site_1: XhoI; Site_2: EcoRI
; The library was made from poly-A RNA in the cloning
vector lambda ZAP II. Clones to be sequenced were
prepared by mass excision."

BASE COUNT 151 a 128 c 173 g 146 t
ORIGIN

Query Match 8.7%; Score 42; DB 10; Length 598;
Best Local Similarity 59.0%; Pred. No. 0.043;
Matches 72; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

QY 60 CGCCAGAACTTCAGCGCCAGGAAGTTGGCGGCCACCGTGGTGTGGCGGTGT 119
DB 279 CGCGAGAGCTGACCGACGAGGAGTTCGACGAGATGATCGTGAGCTGACGTGATGG 338

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 22, 2003, 14:36:28 ; Search time 40.0415 Seconds
(without alignments)
3699.288 Million cell updates/sec

Title: US-09-677-374-3

Perfect score: 483

Sequence: 1 atgcgtggtgcctgcagg.....aggtgattgaccgaaaaa 483

Scoring table: IDENTITY NUC

IDENTITY: NOC
Gapop 10.0 ; Gapext 1.0

Searched: 441362 seqs. 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Maximum Match 1000
Listing first 45 summaries

Database : Issued Patents NA: *

Database : issued_patents_na.
1: /cqn2_6/ptodata/1/ina/5A_COMB.seq:*

```
1: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
```

```
2: /cgn2_6/ptodata/1/ina/6A_COMB.seq;
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq; *
```

4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*

5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*

```
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query | | | DB | ID | Description |
|------------|-------|-------|---------|---|-------------------|--------------------|-------------|
| | | Match | Length | | | | |
| 1 | 36.6 | 7.6 | 303 | 4 | US-08-556-978B-80 | Sequence 80, Appl | |
| 2 | 36.6 | 7.6 | 303 | 4 | US-08-556-978B-81 | Sequence 81, Appl | |
| C 3 | 33.6 | 7.0 | 4403765 | 4 | US-09-103-840A-2 | Sequence 2, Appl | |
| C 4 | 33.6 | 7.0 | 4411529 | 4 | US-09-103-840A-1 | Sequence 1, Appl | |
| C 5 | 33.4 | 6.9 | 962 | 4 | US-08-765-907A-16 | Sequence 16, Appl | |
| C 6 | 33.4 | 6.9 | 1052 | 2 | US-08-403-852D-10 | Sequence 10, Appl | |
| C 7 | 33.4 | 6.9 | 1052 | 3 | US-08-510-646B-10 | Sequence 10, Appl | |
| C 8 | 33.4 | 6.9 | 1052 | 4 | US-09-231-818-10 | Sequence 10, Appl | |
| C 9 | 33.4 | 6.9 | 2888 | 4 | US-08-763-907A-1 | Sequence 1, Appl | |
| 10 | 33.2 | 6.9 | 23673 | 4 | US-09-773-816-1 | Sequence 1, Appl | |
| 11 | 33.2 | 6.9 | 4403765 | 4 | US-09-103-840A-2 | Sequence 2, Appl | |
| 12 | 33.2 | 6.9 | 4411529 | 4 | US-09-103-840A-1 | Sequence 1, Appl | |
| 13 | 33 | 6.8 | 981 | 4 | US-09-307-973A-3 | Sequence 3, Appl | |
| C 14 | 33 | 6.8 | 1295 | 1 | US-08-332-747-1 | Sequence 1, Appl | |
| C 15 | 32.4 | 6.7 | 535 | 6 | 5187077-4 | Patent No. 5187077 | |
| C 16 | 32.4 | 6.7 | 535 | 6 | 5427925-3 | Patent No. 5427925 | |
| C 17 | 32.4 | 6.7 | 599 | 6 | 5187077-24 | Patent No. 5187077 | |
| C 18 | 32.4 | 6.7 | 599 | 6 | 5427925-22 | Patent No. 5427925 | |
| C 19 | 32.4 | 6.7 | 714 | 2 | US-08-378-617A-25 | Sequence 25, Appl | |
| C 20 | 32.4 | 6.7 | 1437 | 6 | 5187077-16 | Patent No. 5187077 | |
| C 21 | 32.4 | 6.7 | 1437 | 6 | 5427925-14 | Patent No. 5427925 | |
| C 22 | 32.4 | 6.7 | 1820 | 4 | US-09-530-212A-1 | Sequence 1, Appl | |
| C 23 | 32 | 6.6 | 1869 | 4 | US-09-372-668-3 | Sequence 3, Appl | |
| C 24 | 32 | 6.6 | 2365 | 1 | US-08-363-208-1 | Sequence 1, Appl | |
| C 25 | 32 | 6.6 | 2365 | 4 | US-09-137-478-1 | Sequence 1, Appl | |
| 26 | 32 | 6.6 | 3300 | 1 | US-08-194-290-6 | Sequence 6, Appl | |
| 27 | 32 | 6.6 | 3300 | 2 | US-08-614-377A-6 | Sequence 6, Appl | |

| | | | | | | |
|----|------|-----|-------|---|--------------------|-------------------|
| 28 | 32 | 6.6 | 3300 | 4 | US-09-142-648B-6 | Sequence 13, Appl |
| 29 | 31.8 | 6.6 | 13987 | 2 | US-08-804-237C-13 | Sequence 13, Appl |
| 30 | 31.8 | 6.6 | 44377 | 2 | US-08-804-227C-7 | Sequence 17, Appl |
| 31 | 31.8 | 6.6 | 44377 | 2 | US-08-804-198-1 | Sequence 1, Appl |
| 32 | 31.6 | 6.5 | 453 | 4 | US-09-397-787-239 | Sequence 239, App |
| 33 | 31.6 | 6.5 | 5036 | 1 | US-09-177-349-2 | Sequence 2, Appl |
| 34 | 31.2 | 6.5 | 7218 | 1 | US-08-232-463-14 | Sequence 14, Appl |
| 35 | 30.8 | 6.4 | 1843 | 1 | US-07-918-023-1 | Sequence 1, Appl |
| 36 | 30.8 | 6.4 | 2171 | 3 | US-08-851-843A-100 | Sequence 100, App |
| 37 | 30.8 | 6.4 | 2171 | 4 | US-08-974-549A-266 | Sequence 266, App |
| 38 | 30.8 | 6.4 | 2171 | 4 | US-08-854-050-100 | Sequence 100, App |
| 39 | 30.8 | 6.4 | 2171 | 4 | US-09-430-333-100 | Sequence 100, App |
| 40 | 30.8 | 6.4 | 2171 | 4 | US-09-408-647A-1 | Sequence 1, Appl |
| 41 | 30.6 | 6.3 | 882 | 4 | US-08-818-112-138 | Sequence 138, App |
| 42 | 30.6 | 6.3 | 882 | 4 | US-08-818-111-133 | Sequence 133, App |
| 43 | 30.6 | 6.3 | 882 | 4 | US-09-056-556-138 | Sequence 138, App |
| 44 | 30.6 | 6.3 | 882 | 4 | US-09-072-592-133 | Sequence 133, App |
| 45 | 30.6 | 6.3 | 1689 | 1 | US-08-442-884-2 | Sequence 2, Appl |

ALIGNMENTS

RESULT 1
US-08-556-378B-80
; Sequence 80, Application US/08556978B
; Patent No. 6268169
; GENERAL INFORMATION:
; APPLICANT: FAHNESTOCK, STEPHEN F.
; TITLE OF INVENTION: NOVEL RECOMBINANTLY PRODUCED
; TITLE OF INVENTION: SPIDER SILK ANALOGS
; NUMBER OF SEQUENCES: 107
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: E. I. DU PONT DE NEMOURS AND COMPANY
; STREET: 1007 MARKET STREET
; CITY: WILMINTON
; STATE: DELAWARE
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 19898
; COMPUTER READABLE FORM:
; MEDIUM TYPE: DISKETTE, 3.50 INCH
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: MICROSOFT WINDOWS 95
; SOFTWARE: MICROSOFT WORD FOR WINDOWS 95
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/556.978B

```

1  FILING DATE:
2  CLASSIFICATION: 435
3  PRIOR APPLICATION DATA:
4  APPLICATION NUMBER: 08/077,600
5  FILING DATE: JUNE 15, 1993
6  ATTORNEY/AGENT INFORMATION:
7  NAME: FLOYD, LINDA AXAMETHY
8  REGISTRATION NUMBER: 33,692
9  REFERENCE/DOCKET NUMBER: CR-9389-A
10 TELECOMMUNICATION INFORMATION:
11 TELEPHONE: 302-892-8112
12 TELEFAX: 302-773-0164
13 INFORMATION FOR SEQ ID NO: 80:
14 SEQUENCE CHARACTERISTICS:
15 LENGTH: 303 base pairs
16 TYPE: nucleic acid
17 STRANDEDNESS: single
18 TOPOLOGY: linear
19 MOLECULE TYPE: DNA (genomic)
20 US-08-556-978B-80

```

| | | | | |
|-----------------------|--------|------------------|-------|-------------|
| Query Match | 7.6% | Score 36.6; | DB 4; | Length 303; |
| Best Local Similarity | 57.4%; | Pred. No. 0.053; | | |

Best local similarity 57.48; Acc. No. 0.055,
Matches 66; Conservative 0; Mismatches 49; Indels 0; Gaps 0;

85 GTTGGCGCGGCCACCGGTGCGGTGTGGCGGTGTGGCGCCAGCTGTTCGGTAAAGC 144

QY 219 GGAC 222
Db 2061290 GGTC 2061287

RESULT 5

US-08-765-907A-16/C
Sequence 16, Application US/08765907A

Patent No. 6352839

GENERAL INFORMATION:

APPLICANT: BLANC, Veronique

APPLICANT: THIBAUT, Denis

APPLICANT: BAWAS-JACQUES, Nathalie

APPLICANT: BLANCHE, Francis

APPLICANT: COUZET, Joel

APPLICANT: BARRIERE, Jean-Claude

APPLICANT: DEBUSSCHE, Laurent

APPLICANT: FAMECHON, Alain

APPLICANT: PARIS, Jean-Marc

APPLICANT: DUTRUC-ROSSET, Gilles

TITLE OF INVENTION: Streptogramins And Method For Preparing Same By

TITLE OF INVENTION: Mutasythesis

FILE REFERENCE: Streptogramin genes

CURRENT APPLICATION NUMBER: US/08/765,907A

CURRENT FILING DATE: 1997-03-20

NUMBER OF SEQ ID NOS: 17

SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 16

LENGTH: 962

TYPE: DNA

ORGANISM: Streptomyces pristinaespiralis

US-08-765-907A-16

Query Match 6.9%; Score 33.4; DB 4; Length 962;

Best Local Similarity 55.7%; Pred. No. 0.82;

Matches 64; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

QY 75 CCGCCAGAGTTCGGCGGCCACCGTGTGCGGTGTGGCGGTGTCGGCGGCGAGCTGTTGCCGCGCAGCTGTT 134

Db 243 CGGCCAGCAGGTTCGAGGTTCCTCCCGCGGCGACCTGGCAGGCGTGGCGCGCCAGCAGCC 184

QY 135 CGGTAAAGGCTCTGCTGTGTCGATGGCCATCGCGGTCGCGTTCGTGGCGGT 189

Db 183 GGGTGTGGCGCGCGCGGTCGCGGTCGATCCCGCGCGCGTTCAGCTGCCCGGT 129

RESULT 6

US-08-403-852D-10/C

Sequence 10, Application US/08403852D

Patent No. 5891695

GENERAL INFORMATION:

APPLICANT: Blanc, Veronique

APPLICANT: Blanche, Francis

APPLICANT: Crouzet, Joel

APPLICANT: Jacques, Nathalie

APPLICANT: Lacroix, Patricia

APPLICANT: Thibaut, Denis

APPLICANT: Zagorec, Monique

APPLICANT: Debussche, Laurent

APPLICANT: De Crecy-Lagard, Valerie

TITLE OF INVENTION: Polypeptides Involved In The

TITLE OF INVENTION: Biosynthesis Of Streptogramins, Nucleotide Sequences

TITLE OF INVENTION: Coding For These Polypeptides And Their Use

NUMBER OF SEQUENCES: 43

CORRESPONDENCE ADDRESS:

ADDRESSEE: Finnegan, Henderson, Farabow, Garrett & Dunner

STREET: 1300 I Street, N.W., Suite 700

CITY: Washington

STATE: D.C.

COUNTRY: USA

ZIP: 20005-3315

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/403,852D
FILING DATE: 10-MAY-1995

PRIOR APPLICATION DATA: PCT/FR 93/00923

FILING DATE: 25-SEP-1993

PRIOR APPLICATION DATA: FR 92/11441

FILING DATE: 25-SEP-1992

ATTORNEY/AGENT INFORMATION:

NAME: Meyers, Kenneth J.

REGISTRATION NUMBER: 25,146

REFERENCE/DOCKET NUMBER: 03806.0054-00000

TELECOMMUNICATION INFORMATION:

TELEPHONE: (202) 408-4000

TELEFAX: (202) 408-4400

INFORMATION FOR SEQ ID NO: 10:

SEQUENCE CHARACTERISTICS:

LENGTH: 1052 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: CDNA

HYPOTHETICAL: NO

ANTI-SENSE: NO

ORIGINAL SOURCE:

ORGANISM: S.pristinaespiralis

FEATURE:

NAME/KEY: CDS

LOCATION: 84..962

OTHER INFORMATION: /product= "Gene PapM"

US-08-403-852D-10

Query Match 6.9%; Score 33.4; DB 2; Length 1052;

Best Local Similarity 55.7%; Pred. No. 0.86;

Matches 64; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

QY 75 CCGCCAGAGTTCGGCGGCCACCGTGTGCGGTGTGGCGGTGTTGGCGGCGAGCTGTT 134

Db 243 CGGCCAGCAGGTTCGAGGTTCCTCCCGCGGCGACCTGGCAGGCGTGGCGCGCCAGCAGCC 184

QY 135 CGGTAAAGGCTCTGCTGTGTCGATGGCCATCGCGGTCGCGTTCGTGGCGGT 189

Db 183 GGGTGTGGCGCGCGCGGTCGCGGTCGATCCCGCGCGCGTTCAGCTGCCCGGT 129

RESULT 7

US-08-510-646B-10/C

Sequence 10, Application US/08510646B

Patent No. 6077699

GENERAL INFORMATION:

APPLICANT: Blanc, Veronique

APPLICANT: Blanche, Francis

APPLICANT: Crouzet, Joel

APPLICANT: Jacques, Nathalie

APPLICANT: Lacroix, Patricia

APPLICANT: Thibaut, Denis

APPLICANT: Zagorec, Monique

APPLICANT: Debussche, Laurent

APPLICANT: De Crecy-Lagard, Valerie

TITLE OF INVENTION: Polypeptides Involved In The

TITLE OF INVENTION: Biosynthesis Of Streptogramins, Nucleotide Sequences

TITLE OF INVENTION: Coding For These Polypeptides And Their Use

NUMBER OF SEQUENCES: 45

CORRESPONDENCE ADDRESS:

ADDRESSEE: Finnegan, Henderson, Farabow, Garrett & Dunner

STREET: 1300 I Street, N.W., Suite 700

CITY: Washington

STATE: D.C.

COUNTRY: USA

ZIP: 20005-3315
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/510.646B
FILING DATE: 03-AUG-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/403,852
FILING DATE: 10-MAY-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/FR 93/00923
FILING DATE: 25-SEP-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: FR 92/11441
FILING DATE: 25-SEP-1992
ATTORNEY/AGENT INFORMATION:
NAME: Meyers, Kenneth J.
REGISTRATION NUMBER: 25,146
REFERENCE/DOCKET NUMBER: 03806.0054-01000
TELEPHONE: (202) 408-4000
TELEFAX: (202) 408-4400
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 1052 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: CDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: S.pristinaespiralis
FEATURE:
NAME/KEY: CDS
LOCATION: 84..962
OTHER INFORMATION: /product= "Gene PapM"
US-08-510-646B-10

Query Match 6.9%; Score 33.4; DB 3; Length 1052;
Best Local Similarity 55.7%; Pred. No. 0.86;
Matches 64; Conservative 0; Mismatches 51; Indels 0; Gaps 0;
QY 75 CGCCAGGAGTGGCGCGCCACCGGTGGTGGCGGTGGTGGCGCGGTGGT 134
DB 243 CGCCAGGAGTGGCGCGCCACCGGTGGTGGCGGTGGTGGCGCGGTGGT 184
QY 135 CGGTAAAGCTCTGTCGTGTCGATGTCATGCCATCGCGGTGGTGGCGGT 189
DB 183 GGTGTGCGCGCGCGCGGTGGTGGCGGTGGTGGCGGTGGTGGCGGT 129

RESULT 8
US-09-231-818-10/c
Sequence 10, Application US/09231818
Patent No. 6171846
GENERAL INFORMATION:
APPLICANT: Blanc, Veronique
APPLICANT: Blanc, Francis
APPLICANT: Crouzet, Joel
APPLICANT: Jacques, Nathalie
APPLICANT: Lacroix, Patricia
APPLICANT: Thibaut, Denis
APPLICANT: Zagorec, Monique
APPLICANT: De Crecy-Lagard, Valerie
TITLE OF INVENTION: Polypeptides Involved In The
Biosynthesis Of Streptogramins, Nucleotide Sequences
Coding For These Polypeptides And Their Use

NUMBER OF SEQUENCES: 43
CORRESPONDENCE ADDRESS:
ADDRESSEE: Finnegan, Henderson, Parabow, Garrett & Dunner
STREET: 1300 I Street, N.W., Suite 700
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20005-3315
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/231.818
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/403,852
FILING DATE: 10-MAY-1995
APPLICATION NUMBER: PCT/FR 93/00923
FILING DATE: 25-SEP-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: FR 92/11441
FILING DATE: 25-SEP-1992
ATTORNEY/AGENT INFORMATION:
NAME: Meyers, Kenneth J.
REGISTRATION NUMBER: 25,146
REFERENCE/DOCKET NUMBER: 03806.0054-00000
TELEPHONE: (202) 408-4000
TELEFAX: (202) 408-4400
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 1052 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: CDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: S.pristinaespiralis
FEATURE:
NAME/KEY: CDS
LOCATION: 84..962
OTHER INFORMATION: /product= "Gene PapM"
US-09-231-818-10

Query Match 6.9%; Score 33.4; DB 4; Length 1052;
Best Local Similarity 55.7%; Pred. No. 0.86;
Matches 64; Conservative 0; Mismatches 51; Indels 0; Gaps 0;
QY 75 CGCCAGGAGTGGCGCGCCACCGGTGGTGGCGGTGGTGGCGCGGTGGT 134
DB 243 CGCCAGGAGTGGCGCGCCACCGGTGGTGGCGGTGGTGGCGCGGTGGT 184
QY 135 CGGTAAAGCTCTGTCGTGTCGATGTCATGCCATCGCGGTGGTGGCGGT 189
DB 183 GGTGTGCGCGCGCGCGGTGGTGGCGGTGGTGGCGGTGGTGGCGGT 129

RESULT 9
US-08-765-907A-1/c
Sequence 1, Application US/08765907A
Patent No. 6352839
GENERAL INFORMATION:
APPLICANT: Blanc, Veronique
APPLICANT: Thibaut, Denis
APPLICANT: Bamas-Jacques, Nathalie
APPLICANT: Blanche, Francis
APPLICANT: Couzet, Joel
APPLICANT: Barriere, Jean-Claude
APPLICANT: Debuysche, Laurent

| | | | |
|----------------|-----|---|-----|
| Qy | 37 | A T C T G T F T T T C T G T T G G T T G C G C C A G A A C T T C A G C C G C A G A A G T T G S C G C G C C | 96 |
| | | | |
| D _b | 655 | A T C T G S G C G G C A A C T G G G T T T A C C C T G T G E T T C T G S G C A A T G C T G T G T G S C A G C G | 714 |
| | | | |
| Qy | 97 | A C G G T G C G G T T T G G G C G G T G T T G C G G C C A C T G T C G G T A A A G G C T C T G G T C G T G T G | 156 |
| | | | |
| D _b | 715 | A C G G G T G G A T G G T G G C G T C A G T G G C G C T G G C G G T G C T A T C G G C T T T A T T G T C T G | 774 |
| Qy | 157 | T C G A T G G C C | 165 |
| | | | |
| D _b | 775 | G T G A T C C C | 783 |

RESULT 14
 US-08-332-747-1/c
 ; Sequence 1, Application US/08332747
 ; Patent No. 5733724
 ;
 ; GENERAL INFORMATION:
 ; APPLICANT: Bubert, Andreas
 ; APPLICANT: Goebel, Werner
 ; APPLICANT: Goetz, Monika
 ; APPLICANT: Ludwig, Albrecht
 ; APPLICANT: Schubert, Peter
 ; APPLICANT: Neumann, Siegfried
 ; TITLE OF INVENTION: Oligonucleotides for the detection of
 ; TITLE OF INVENTION: Enterobacteriaceae
 ; NUMBER OF SEQUENCES: 26

*
*, CORRESPONDENCE ADDRESS:
*, ADDRESSEE: MILLEN, WHITE, ZELANO & BRANIGAN
*, STREET: 2200 CLARENDON BLVD., SUITE 1400
*, CITY: ARLINGTON
*, STATE: VA
*, COUNTRY: USA
*, ZIP: 22201
*, COMPUTER READABLE FORM:
*, MEDIUM TYPE: Floppy disk
*, COMPUTER: IBM PC Compatible
*, OPERATING SYSTEM: PC-DOS/MS-DOS
*, SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
*, CURRENT APPLICATION DATA:
*, APPLICATION NUMBER: US/08/332,747
*, FILING DATE: 01-NOV-1994
*,

```

, PRIOR APPLICATION DATA:
, APPLICATION NUMBER: DE 43 37 295
, FILING DATE: 02-NOV-1993
, ATTORNEY/AGENT INFORMATION:
, NAME: Hamlet-King, Diana
, REGISTRATION NUMBER: 33,302
, REFERENCE/DOCKET NUMBER: MERCK 1628
, TELECOMMUNICATION INFORMATION:
, TELEPHONE: 703-243-6333
, TELEFAX: 703 243-6410
, TELEX: 64191
, INFORMATION FOR SEQ ID NO: 1:
, SEQUENCE CHARACTERISTICS:
, LENGTH: 1295 base pairs
, TYPE: nucleic acid
, STRANDEDNESS: double
, TOPOLOGY: linear
, MOLECULE TYPE: DNA (genomic)
, HYPOTHETICAL: NO
, ANTI-SENSE: NO
, ORIGINAL SOURCE:
, ORGANISM: Salmonella sp.
, US-08-332-747-1

```

| Query Match | 6.8% | Score 33; | DB 1; | Length 1295; |
|-----------------------|-----------------|-----------------------|--|--------------|
| Best Local Similarity | 55.8%; | Pred. No. 1.2; | | |
| Matches 63; | Conservative 0; | Mismatches 50; | Indels 0; | Gaps 0; |
| QY | 86 | TTGGCGGCGCACGGTGTCGGT | TTGTGGCGGCGTGTGGCGGCACGCTGTCGGTAAAGSCT | 145 |
| | | | | |

Db 942 TTGGCGCTATCGGGCGCGGTGTGGCGGATTCCTCGCAACACTATCGGCGGGTA 883
QY 146 CTGGTCGTGTGTCGATGGCCATCGGGGTCGGTTCCTGGCGGCTCTGATTGGC 198
Db 882 CGGGTCGTTCACTGGCGACCGCGGCGGCTGGTGGCTGGCGGTAGCGGGC 830

RESULT 15

5187077-4/c
; Patent No. 5187077
; APPLICANT: GEARING, DAVID P.; GOUGH, NICHOLAS M.; HILTON,
; DOUGLAS J.; KING, JULIE A.; METCALF, DONALD; NICE, EDOUARD C.
; NICOLA, NICOS A.; SIMPSON, RICHARD J.; WILLSON, TRACY A.
; TITLE OF INVENTION: LEUKEMIA INHIBITORY FACTOR
; NUMBER OF SEQUENCES: 41
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/294,514
; FILING DATE: 09-DEC-1988
; SEQ ID NO: 4;
; LENGTH: 535
5187077-4

Query Match 6.7%; Score 32.4; DB 6; Length 535;
Best Local Similarity 47.9%; Pred. No. 1.3;
Matches 93; Conservative 0; Mismatches 101; Indels 0; Gaps 0;
QY 32 TCATTATCTCTGTTTCTCTGGTGGTGGCCCGAGAACITTCAGCGGCCAGGAAGTTGGCG 91
Db 482 TCTTTGTACAGTGGTGGGGGACAGGTGGCACATCCACGTGGCCACACCGTACTTGTG 423
QY 92 CGGCCACCGGTGGGTGGTGGCGGTGTTCCCGCCAGCTGTTCCGTTAAAGGCTCTGGTC 151
Db 422 CACAGAGCGGCAAGACACATTGCTGAGGAGGCCCTCATGACGTCTATAGTAGCATTGAGC 363
QY 152 GTGTGTGATGCCATCGCGGTGGGTCTGGCGGTCTGATTGGCTCTAAATCGGTC 211
Db 362 TTGACCTGGAGGCTCAGCGGAGTGGGTTTCAGACCTTCTGGTCCCGGTGATATGGTC 303
QY 212 AGAGCATGGACCAG 225
Db 302 AGGAGGCGCTCAG 289

Search completed: February 22, 2003, 18:48:18
Job time : 6109.04 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.4_p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 22, 2003, 16:18:29 ; Search time 46.9931 Seconds
(without alignments)
5772.590 Million cell updates/sec

Title: US-09-677-374-3
Perfect score: 483
Sequence: 1 atgcgtggtgcctgcagg.....aggtattagaccgaaaaa 483

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 442118 seqs, 280819700 residues

Total number of hits satisfying chosen parameters: 884236

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications_NA.*
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq.*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq.*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq.*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq.*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq.*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq.*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq.*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq.*
9: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq.*
10: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq.*
11: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq.*
12: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq.*
13: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq.*
14: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|--------------------|
| 1 | 36 | 7.5 | 7791 | 10 | US-09-764-877-3700 |
| 2 | 35.6 | 7.4 | 2481 | 10 | US-09-729-674-141 |
| 3 | 34.8 | 7.2 | 426 | 9 | US-10-125-815-6 |
| 4 | 33.8 | 7.0 | 810 | 9 | US-09-738-626-1633 |
| 5 | 33.6 | 7.0 | 1164 | 9 | US-09-738-626-2633 |
| 6 | 33.2 | 6.9 | 1047 | 10 | US-09-780-996-6 |
| 7 | 33.2 | 6.9 | 1278 | 9 | US-10-086-738A-1 |
| 8 | 33.2 | 6.9 | 1792 | 10 | US-09-822-830A-420 |
| 9 | 32.8 | 6.8 | 900 | 9 | US-09-712-363-88 |
| 10 | 32 | 6.6 | 1869 | 9 | US-10-115-195-3 |
| 11 | 32 | 6.6 | 2364 | 9 | US-10-057-6098-1 |
| 12 | 32 | 6.6 | 2863 | 9 | US-09-822-846-108 |
| 13 | 32 | 6.6 | 3300 | 10 | US-09-379-931-6 |
| 14 | 31.8 | 6.6 | 891 | 9 | US-09-738-626-2169 |
| 15 | 31.8 | 6.6 | 1350 | 10 | US-09-963-690-1 |
| 16 | 31.6 | 6.5 | 453 | 10 | US-09-876-889-239 |
| 17 | 31.6 | 6.5 | 5036 | 10 | US-09-918-951-2 |
| 18 | 31.4 | 6.5 | 3919 | 10 | US-09-070-927A-681 |
| 19 | 31.2 | 6.5 | 1204 | 10 | US-09-822-849A-22 |

| | | | | | | |
|------|------|-----|-------|----|---------------------|-------------------|
| 20 | 31.2 | 6.5 | 2678 | 10 | US-09-964-824A-252 | Sequence 252, App |
| c 21 | 31.2 | 6.5 | 4397 | 10 | US-09-728-952-43 | Sequence 43, Appl |
| c 22 | 31.2 | 6.5 | 4397 | 10 | US-09-728-952-98 | Sequence 98, Appl |
| c 23 | 31 | 6.4 | 116 | 9 | US-10-025-380-856 | Sequence 856, App |
| c 24 | 31 | 6.4 | 116 | 10 | US-09-922-217-856 | Sequence 856, App |
| c 25 | 31 | 6.4 | 116 | 10 | US-09-833-263-856 | Sequence 856, App |
| c 26 | 31 | 6.4 | 147 | 10 | US-09-974-300-3752 | Sequence 3732, Ap |
| 27 | 30.8 | 6.4 | 2171 | 9 | US-09-843-676-100 | Sequence 100, App |
| 28 | 30.8 | 6.4 | 2171 | 9 | US-09-766-253-100 | Sequence 100, App |
| 29 | 30.8 | 6.4 | 2171 | 9 | US-09-438-486-100 | Sequence 100, App |
| c 30 | 30.8 | 6.4 | 2171 | 9 | US-10-053-758-100 | Sequence 1, Appl |
| c 31 | 30.6 | 6.3 | 4030 | 9 | US-09-950-046-1 | Sequence 292, App |
| c 32 | 30.6 | 6.3 | 43058 | 10 | US-09-954-456-292 | Sequence 529, App |
| c 33 | 30.6 | 6.3 | 43058 | 10 | US-09-954-456-529 | Sequence 350, Ap |
| c 34 | 30.6 | 6.3 | 43058 | 10 | US-09-880-107-3950 | Sequence 27332, A |
| c 35 | 30.4 | 6.3 | 411 | 10 | US-09-864-761-27332 | Sequence 10686, A |
| c 36 | 30.4 | 6.3 | 492 | 10 | US-09-864-761-10686 | Sequence 105, App |
| c 37 | 30.4 | 6.3 | 1317 | 9 | US-09-712-363-105 | Sequence 12054, A |
| 38 | 30.2 | 6.3 | 509 | 10 | US-09-864-761-12054 | Sequence 8, Appl |
| 39 | 30.2 | 6.3 | 1035 | 9 | US-10-034-158-8 | Sequence 12, Appl |
| 40 | 30.2 | 6.3 | 1035 | 10 | US-09-086-118-28 | Sequence 12, Appl |
| 41 | 30.2 | 6.3 | 1035 | 10 | US-09-835-684-12 | Sequence 12, Appl |
| 42 | 30.2 | 6.3 | 1035 | 10 | US-09-880-371-12 | Sequence 16, Appl |
| 43 | 30.2 | 6.3 | 1035 | 10 | US-09-879-248-16 | Sequence 8, Appl |
| 44 | 30.2 | 6.3 | 1035 | 10 | US-09-770-693-8 | Sequence 8, Appl |
| 45 | 30.2 | 6.3 | 1035 | 10 | US-09-766-348-8 | Sequence 8, Appl |

ALIGNMENTS

RESULT 1
US-09-764-877-3700
; Sequence 3700, Application US/09764877
; Patent No. US20020147140A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC005
; CURRENT APPLICATION NUMBER: US/09/764.877
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 4031
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3700
; LENGTH: 7791
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-877-3700

| | | | | |
|-----------------------|-------|--|--------|----------------------------------|
| Query Match | 7.5% | Score 36; | DB 10; | Length 7791; |
| Best Local Similarity | 51.2% | Pred. No. 0.35; | | |
| Matches | 84; | Conservative | 0; | Mismatches 80; Indels 0; Gaps 0; |
| QY | 35 | TTATCTCTGTTTCTGGTGGTGGCCGAGAACTTCAGCCGCGGAGAACTTGGCGGG | 94 | |
| Db | 1025 | TTCCCTTTCTTGGCTTGGGTCGCGCGGCGGCGGCGGCGGCGGCGGCGG | 1084 | |
| QY | 95 | CCACCGGTGGGTGGTGGCGGTGTGCGGCCAGCTTTCGTAAGGCTCTGTCGTG | 154 | |
| Db | 1085 | CCCCGGGGGGGGAAGGGGCGGCTGGAGGCTGTGCAGCGAGGGCGGCGGAG | 1144 | |
| QY | 155 | TGTCGATGGCCATCGCGGTGGCTTCTGGCGGCTCTGATTGGC | 198 | |
| Db | 1145 | GGAGGAGCGGCTCGCGGGAGGTGGAGCAGGCTCAGAGCC | 1188 | |

RESULT 2
US-09-729-674-141/c
; Sequence 141, Application US/09729674
; Patent No. US2001003935A1
; GENERAL INFORMATION:
; APPLICANT: Jacobs, Kenneth

Db 894 GTTTGGCAGAGCTGAATGTCTGTCTCCAGCAGCCAGCCATATGCTGTGTCAGTC 953
 QY 417 ACAGAAATCTACGGCACCGCGTGCCTCA 446
 Db 954 ATACGGGTGGAAGACGGGTCTTGCAGTTA 983
 RESULT 7
 US-10-086-738A-1
 ; Sequence 1, Application US/10086738A
 ; Publication No. US20030004306A1
 ; GENERAL INFORMATION:
 ; APPLICANT: MIYAWAKI, Atsushi
 ; APPLICANT: NAGAI, Takeharu
 ; TITLE OF INVENTION: A Fluorescent Protein
 ; FILE REFERENCE: P22042
 ; CURRENT APPLICATION NUMBER: US/10/086,738A
 ; CURRENT FILING DATE: 2002-07-12
 ; PRIOR APPLICATION NUMBER: PCT/JP05/98922001
 ; PRIOR FILING DATE: 2001-03-05
 ; NUMBER OF SEQ ID NOS: 5
 ; SOFTWARE: PatentIn version 3.0
 ; SEQ ID NO 1
 ; LENGTH: 1278
 ; TYPE: DNA
 ; ORGANISM: Artificial Sequence
 ; FEATURE:
 ; OTHER INFORMATION: Construct from several species
 US-10-086-738A-1
 Query Match 6.9%; Score 33.2; DB 9; Length 1278;
 Best Local Similarity 55.1%; Pred. No. 1.1;
 Matches 65; Conservative 0; Mismatches 53; Indels 0; Gaps
 QY 337 TACCAGCGTTACACAAACAGGACCGCTCAGCAGTACTGCCCGGAATTTTCAGCAGAA 396
 Db 733 TTCGAGGCGCACACCCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGAC 792
 QY 397 GCCATGATCGCAGGTGCAGAACAGAGAAATCTACGGCACCGCGTGCCTCAGCCGGATG 454
 Db 793 GGCACATCTCGGGGCACACAGCTGGATACACGCTACCGTACCGGGNCCAACTGACAGAAG 850
 RESULT 8
 US-09-822-830A-420
 ; Sequence 420, Application US/09822830A
 ; Patent No. US20020142952A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Genetics Institute, Inc.
 ; APPLICANT: Wong, Gordon G.
 ; APPLICANT: Clark, Hilary
 ; APPLICANT: Fechtel, Kim
 ; APPLICANT: Agostino, Michael J.
 ; APPLICANT: Howes, Steven H.
 ; APPLICANT: Resnick, Richard J.
 ; APPLICANT: Gulukota, Kamalakara
 ; APPLICANT: Graham, James R.
 ; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING NOVEL SECRETED PROTEINS
 ; FILE REFERENCE: GIN 6402
 ; CURRENT APPLICATION NUMBER: US/09/822,830A
 ; CURRENT FILING DATE: 2001-03-29
 ; PRIOR APPLICATION NUMBER: 60/195,604
 ; PRIOR FILING DATE: 2000-04-06
 ; NUMBER OF SEQ ID NOS: 631
 ; SOFTWARE: PatentIn ver. 2.0
 ; SEQ ID NO 420
 ; LENGTH: 1792
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 US-09-822-830A-420
 Query Match 6.9%; Score 33.2; DB 10; Length 1792;
 Best Local Similarity 51.3%; Pred. No. 1.3;

Db 293 ATGAGCAAGAGATAGAGAAACAGA 270

RESULT 12

```

US-09-822-846-108/c
; Sequence 108, Application US/09822846
; Publication No. US20030037139A1
; GENERAL INFORMATION:
; APPLICANT: Jacobs, Kenneth
; APPLICANT: McCoy, John M.
; APPLICANT: LaVallie, Edward R.
; APPLICANT: Collins-Racie, Lisa A.
; APPLICANT: Evans, Cheryl
; APPLICANT: Merberg, David
; APPLICANT: Treacy, Maurice
; APPLICANT: Agostino, Michael J.
; APPLICANT: Steininger II, Robert J.
; APPLICANT: Bowman, Michael R.
; APPLICANT: Spaulding, Vikki
; APPLICANT: Wong, Gordon G.
; APPLICANT: Clark, Hilary
; APPLICANT: Fechtel, Kim
; APPLICANT: Howes, Steven H.
; APPLICANT: Resnick, Richard J.
; APPLICANT: Gulukota, Kamalakara
; APPLICANT: Graham, James R.
; APPLICANT: Genetics Institute, Inc.
; TITLE OF INVENTION: POLYNUCLEOTIDES 1
; FILE REFERENCE: GIN 6400
; CURRENT APPLICATION NUMBER: US/09/822
; CURRENT FILING DATE: 2001-03-29
; PRIOR APPLICATION NUMBER: 60/195,605
; PRIOR FILING DATE: 2000-04-06
; NUMBER OF SEQ ID NOS: 629
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 108
; LENGTH: 2863
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-822-846-108

```

| | Query Match | 6.6% | Score 32; | DB 9; | Length 2863; |
|----|-----------------------|--|----------------|-----------|--------------|
| | Best Local Similarity | 51.4%; | Pred. No. 4; | | |
| | Matches 74; | Conservative 0; | Mismatches 70; | Indels 0; | Gaps 0; |
| Qy | 6 | TGGTTGGCTGCAGGGCAGCTCTCTGATCATTTATCTCTGTTTCTTGGTGGGTGGGCCCA | 65 | | |
| | | | | | |
| Db | 2345 | TGTTTGGCTGCAGGGCTCGACTGGGGGGGTGTCTCTGGGCGCAGGTGGGGGCTGGGGCCA | 2186 | | |
| Qy | 66 | GAACCTTCAGCCGCCAGGAAGTTGGCGCGGCCACCGGTGCGGTGTGTGGCGCGTGTGGCGGG | 125 | | |
| | | | | | |
| Db | 2185 | GGACCGGGGCCCTCTGACAGCCCTTGGGCGAAGGATATGATGGGGAGGGGTGGCTG | 2126 | | |
| Qy | 126 | CCAGCTGTTCGTTAAAGGCTCTGG | 149 | | |
| | | | | | |
| Db | 2125 | CCACGGGGGAACAGGGGCCCTGG | 2102 | | |

RECEIVED 13

RES001 13
US-09-379-931-6
; Sequence 6, Application US/09379931
; Patent No. US20020009792A1
; GENERAL INFORMATION:
; APPLICANT: Smit, John
; APPLICANT: Bingle, Wade H.
; APPLICANT: NO. US20020009792A1ellini, John F.
; TITLE OF INVENTION: EXPRESSION AND SECRETION OF HETEROLOGOUS POLYPEPTIDES FROM CAULOC
; FILE REFERENCE: 08106/002003
; CURRENT APPLICATION NUMBER: US/09/379,931
; CURRENT FILING DATE: 1999-08-24
; PRIOR APPLICATION NUMBER: US 08/614,377
; PRIOR FILING DATE: 1996-03-12

```

; PRIOR APPLICATION NUMBER: US 08/194,290
; PRIOR FILING DATE: 1994-02-09
; PRIOR APPLICATION NUMBER: US 07/895,367
; PRIOR FILING DATE: 1992-06-09
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO: 6
; LENGTH: 3300
; TYPE: DNA
; ORGANISM: Caulobacter crescentus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (101)...(3178)
; US-09-379,931-6

```

Query Match

Query Match 0.0%; Score 32; DB 10; Length 3300;
Best Local Similarity 48.4%; Pred. No. 4.3;
Matches 89; Conservative 0; Mismatches 95; Indels 0; Gaps 0;

| | | | |
|----|------|--|------|
| Qy | 24 | CTCTCTGATCATATACTCCTGTTTTCTGTGGTGCGCCAGAACCTTCAGCGGCCAAGA | 83 |
| | | | |
| Db | 856 | CTCGCTGACCACCGGCACCGACACCTCTGACGGGCACGCCAACACGACACACTGCTCCTGC | 915 |
| | | | |
| Qy | 84 | AGTTGGCGCGCCACACGGTCCGGTTGTGGCGCGTGTTCGGGCCACGCTGTTCCGTTAAAGG | 143 |
| | | | |
| Db | 916 | GGGTGAAGTCGCGCGGCGCTCCGACCTGACCGTTGGCGCACACCTGAGCGGGGCTGCTGG | 975 |
| | | | |
| Qy | 144 | CTCTGGTCTGTGTCTGATGGCCATCGCGCGTTCGGGTCTTCTGGCGGCTCATTTGGCTCTAA | 203 |
| | | | |
| Db | 976 | CACCGACGTCCTGAATGGGTGCAAGCTGCTGCGGTTACGGCTCTCCGACCGGCGTGAC | 1031 |
| | | | |
| Qy | 204 | AATC | 207 |
| | | | |
| Db | 1036 | GATC | 1039 |

RESULT 14

```

US-09-738-626-2169/c
: Sequence 2169, Application US/09738626
: Publication No. US20002197605A1
: GENERAL INFORMATION:
: APPLICANT: NAKAGAWA, SATOSHI
: APPLICANT: MIZOGUCHI, HIROSHI
: APPLICANT: ANDO, SEIKO
: APPLICANT: HAYASHI, MIKIRO
: APPLICANT: OCHIAI, KEIKO
: APPLICANT: YOKAI, HARUHIKO
: APPLICANT: TATEISHI, NAOKO
: APPLICANT: SENOH, AKIHIRO
: APPLICANT: IKEDA, MASATO
: APPLICANT: OZAKI, AKIO
: TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
: FILE REFERENCE: 249-1325
: CURRENT APPLICATION NUMBER: US/09/738,626
: CURRENT FILING DATE: 2000-12-18
: PRIOR APPLICATION NUMBER: JP 99/377484
: PRIOR FILING DATE: 1999-12-16
: PRIOR APPLICATION NUMBER: JP 00/159162
: PRIOR FILING DATE: 2000-04-07
: PRIOR APPLICATION NUMBER: JP 00/280988
: PRIOR FILING DATE: 2000-08-03
: NUMBER OF SEQ ID NOS: 7059
: SOFTWARE: PatentIn ver. 3.0
: SEQ ID NO 2169
: LENGTH: 891
: TYPE: DNA
: ORGANISM: Corynebacterium glutamicum
US-09-738-626-2169

```

```
Query Match      6.6%; Score 31.8; DB 9; Length 891;
Best Local Similarity 59.3%; Pred. NO.2.7;
Matches 54; Conservative 0; Mismatches 37; Indels 0; Gaps 0;
```


GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 22, 2003, 13:11:08 ; Search time 2141.73 Seconds
(without alignments)
10435.915 Million cell updates/sec

Title: US-09-677-374-5

Perfect score: 768

Sequence: 1 atgtcagtgaattctacaa.....agggtattgaccgcgaaaaa 768

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl:*

1: gb_ba:*

2: gb_hgt:*

3: gb_in:*

4: gb_om:*

5: gb_ov:*

6: gb_pat:*

7: gb_ph:*

8: gb_pl:*

9: gb_pr:*

10: gb_ro:*

11: gb_sts:*

12: gb_sy:*

13: gb_un:*

14: gb_vl:*

15: em_ba:*

16: em_fun:*

17: em_hum:*

18: em_in:*

19: em_mu:*

20: em_om:*

21: em_or:*

22: em_ov:*

23: em_pat:*

24: em_ph:*

25: em_pl:*

26: em_ro:*

27: em_sts:*

28: em_un:*

29: em_vi:*

30: em_hgt_hum:*

31: em_hgt_inv:*

32: em_hgt_other:*

33: em_hgt_mus:*

34: em_hgt_pln:*

35: em_hgt_rod:*

36: em_hgt_mam:*

37: em_hgt_vrt:*

38: em_sy:*

39: em_hgtgo_hum:*

40: em_hgtgo_mus:*

41: em_hgtgo_other:*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match % | Length | DB | ID | Description |
|------------|-------|---------------|--------|----|------------|---------------------|
| 1 | 295 | 38.4 | 489 | 6 | AX252413 | Sequence AX252413 |
| 2 | 295 | 38.4 | 4983 | 1 | AF184152 | Sequence AF184152 |
| 3 | 270.2 | 35.2 | 486 | 6 | AR001082 | Sequence AR001082 |
| c 4 | 270.2 | 35.2 | 486 | 6 | AR001083 | Sequence AR001083 |
| 5 | 270.2 | 35.2 | 486 | 6 | AR027448 | Sequence AR027448 |
| c 6 | 270.2 | 35.2 | 486 | 6 | AR027449 | Sequence AR027449 |
| 7 | 270.2 | 35.2 | 486 | 6 | AR058257 | Sequence AR058257 |
| c 8 | 270.2 | 35.2 | 486 | 6 | AR058258 | Sequence AR058258 |
| 9 | 270.2 | 35.2 | 486 | 6 | 118514 | Sequence 118514 |
| c 10 | 270.2 | 35.2 | 486 | 6 | 118515 | Sequence 118515 |
| 11 | 270.2 | 35.2 | 486 | 6 | 166614 | Sequence 166614 |
| c 12 | 270.2 | 35.2 | 486 | 6 | 166615 | Sequence 166615 |
| 13 | 270.2 | 35.2 | 486 | 6 | 188789 | Sequence 188789 |
| c 14 | 270.2 | 35.2 | 486 | 6 | 188790 | Sequence 188790 |
| 15 | 270.2 | 35.2 | 499 | 6 | AR096204 | Sequence AR096204 |
| c 16 | 270.2 | 35.2 | 1146 | 6 | AR151733 | Sequence AR151733 |
| 17 | 270.2 | 35.2 | 1146 | 6 | AX250579 | Sequence AX250579 |
| 18 | 270.2 | 35.2 | 6760 | 1 | CLOCBP | Sequence CLOCBP |
| 19 | 76.4 | 9.9 | 6240 | 1 | AB004845 | Sequence AB004845 |
| 20 | 70.8 | 9.2 | 4919 | 1 | CCU40345 | Sequence CCU40345 |
| 21 | 70.2 | 9.1 | 1069 | 12 | AF283517 | Sequence AF283517 |
| 22 | 70.2 | 9.1 | 1482 | 6 | AR129965 | Sequence AR129965 |
| 23 | 70.2 | 9.1 | 1482 | 6 | AR205057 | Sequence AR205057 |
| 24 | 70.2 | 9.1 | 2400 | 1 | CTCIPBA | Sequence CTCIPBA |
| 25 | 70.2 | 9.1 | 5573 | 1 | CLOCIPA | Sequence CLOCIPA |
| 26 | 66.8 | 8.7 | 10898 | 1 | AE007606 | Sequence AE007606 |
| 27 | 61.6 | 8.0 | 6555 | 1 | AF155197 | Sequence AF155197 |
| 28 | 51 | 6.6 | 3048 | 1 | CLOPEEGASE | Sequence CLOPEEGASE |
| 29 | 51 | 6.6 | 10782 | 1 | CTH275974 | Sequence CTH275974 |
| 30 | 47.4 | 6.2 | 3360 | 1 | CSCELZ | Sequence CSCELZ |
| 31 | 46.6 | 6.1 | 537 | 1 | RIR17KGA | Sequence RIR17KGA |
| 32 | 46 | 6.0 | 3067 | 1 | CTCHIAGEN | Sequence CTCHIAGEN |
| c 33 | 45.8 | 6.0 | 8246 | 6 | AX345104 | Sequence AX345104 |
| 34 | 45 | 5.9 | 3120 | 1 | CSCELYZ | Sequence CSCELYZ |
| 35 | 45 | 5.9 | 237523 | 1 | RPXX04 | Sequence RPXX04 |
| c 36 | 44.6 | 5.8 | 147192 | 2 | AC116925 | Sequence AC116925 |
| 37 | 44.4 | 5.8 | 1011 | 1 | CTH420770 | Sequence CTH420770 |
| c 38 | 44.4 | 5.8 | 10595 | 6 | AX251295 | Sequence AX251295 |
| 39 | 44 | 5.7 | 3498 | 4 | AY044905 | Sequence AY044905 |
| 40 | 44 | 5.7 | 3632 | 6 | AX082874 | Sequence AX082874 |
| c 41 | 44 | 5.7 | 48839 | 2 | AC095545 | Sequence AC095545 |
| c 42 | 44 | 5.7 | 125020 | 9 | AF429315 | Sequence AF429315 |
| 43 | 44 | 5.7 | 174256 | 2 | AC107099 | Sequence AC107099 |
| c 44 | 43.8 | 5.7 | 186622 | 2 | AC023433 | Sequence AC023433 |
| 45 | 43.8 | 5.7 | 197164 | 9 | AC064865 | Sequence AC064865 |

ALIGNMENTS

RESULT 1

AX252413

LOCUS

DEFINITION

AX252413

ACCESSION

VERSION

AX252413.1

KEYWORDS

SOURCE

ORGANISM

Piscirickettsia salmonis.

Piscirickettsia salmonis

Bacteria; Proteobacteria; gamma subdivision; Piscirickettsia group;

Piscirickettsia.

1 (bases 1 to 489)

REFERENCE

AUTHORS

Simard,N., Brouwers,H., Jones,S.F., Griffiths,S., Valenzuela,P. and

Burzio,L.

AX252413 Sequence 5 from Patent WO0168865.

AX252413 GI:15985721

AX252413.1

489 bp DNA linear PAT 05-OCT-2001

Fish vaccine against piscirickettsia salmonis


```
CDS
/gene="tnpA"
complement(3842..4561)
/genes="tnpA"
/codon_start=1
/transl_table=11
/product="transposase"
/protein_id="AAG16999.1"
/db_xref="GI:1041345"
/translation="MDTLISLFCVIDDFCQELLFQWNAILEDNKKKNSQSMSTS
EIMTMIYPHKSNYRNFMYLHVIGKSVKYPNSVSYNREVELMPSLLPLCFPIA
AQGKATGYFDYDSTILRYCHEKRSQNRFRGLAKRSKMTGWTYGFKLHIIVNDMG
ELMAFKMSKATYDDRVLPKMAENLTGKIIGDKYISQKLFQDLYEKGQLQITKIRK
MKNKLVLMDKILLKRGIIA"
repeat_region
4562..4849
BASE COUNT 1362 a 778 c 1265 g 1578 t
ORIGIN
Query Match 38.4%; Score 295; DB 1; Length 4983;
Best Local Similarity 76.0%; Pred. No. 1.3e-62;
Matches 364; Conservative 0; Mismatches 115; Indels 0; Gaps 0;
QY 290 GTGGTCCGTCGAGGCGCTCTCTGATCAATATATCTGTGTTTCCTCGGTGGGCGCC 349
DB 2841 GAGGATGTTTCAAGGTAGTAGTCTAATATTATCATGCTGTGTTTATGTTGGCTG 2900
QY 350 AGAAGTTCAGCCCGGAGGTTGGCCCGCCCGCCCGGTCGGTGTGGCGGTGTTGCCG 409
DB 2901 AGAAGTTCAGGAGGTCGAGCTCGGAGCTGGGCGTGTGTTGGCGGTGTTGCTG 2960
QY 410 GCCAGCTGTTTCGCTAAAGGCTCTGCTGTCGATGGCCATCGCGGTTCCTCGG 469
DB 2961 GCCAGCTGTTTGAAGGTAGTGGTCGAGTTCGATGGCCATGGTGGTGTGTTTGG 3020
QY 470 GCGGTCGATGGCTCTAAATCGGTCAGAGCATGGACGAGGATGATGATGATGATG 529
DB 3021 GTGGATTAATGTTCTTAAATCGGTCATCGATGGATGATGATGATGATGATGAT 3080
QY 530 ACCAGTCTCTGGAAGGTTAAAGGTCGACCTACCGGTCGATGATGATGATGATG 589
DB 3081 ACCAGGATTTGGAAGGTTAAAGGTCGACCTACCGGTCGATGATGATGATGATG 3140
QY 590 GTACAGCTACTCTGGAAGGTCGACCTACCGGTCGATGATGATGATGATGATG 649
DB 3141 GCAATAGTATAGTGTGACCGAGTGGTACTTACCGGTCGATGATGATGATGAT 3200
QY 650 GTCAGCAGTACTCCCGGAAATTTTCAGCAGAAAGCCATGATGCGAGGTGATGATG 709
DB 3201 GCCAGCAATATTGTCGAGAAATTTTCAGCAAAAGCGGATGATGCGAGGCGAG 3260
QY 710 TCTACGCGACCGGTCGCTTCAGCGGATGCGCGGTCGCGGTCGATGATGATGAT 768
DB 3261 TTTACGCGACTGTCGCGGCAACCGGATGCGTGTGGCAAGTCATTTTCAACGAA 3319
RESULT 3
AR001082
LOCUS AR001082 486 bp DNA linear PAT 04-DEC-1998
DEFINITION Sequence 1 from patent US 5738984.
ACCESSION AR001082
VERSION AR001082.1 GI:3963149
KEYWORDS
SOURCE
ORGANISM
REFERENCE 1 (bases 1 to 486)
AUTHORS Shoseyov, O.
TITLE Kits and methods of detection using cellulose binding domain fusion proteins
JOURNAL Patent: US 5738984-A 1 14-APR-1998;
FEATURES
Location/Qualifiers
source 194 a 89 c 79 g 124 t
BASE COUNT 194 a 89 c 79 g 124 t
Query Match 35.2%; Score 270.2; DB 6; Length 486;
Best Local Similarity 98.9%; Pred. No. 1.5e-56;
Matches 272; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 ATGTCAGTTGAATCTCAAACTCTAACAATCAGCACAACAACTCAATTACACCAATA 60
DB 471 ATGTCAGTTGAATTTTACAATCTAACAATCAGCACAACAACTCAATTACACCAATA 412
QY 61 ATCAAAATTACTAACACATCTGACAGTGATTTAAATTTAAATGACGTAAAGTTAGATAT 120
DB 411 ATCAAAATTACTAACACATCTGACAGTGATTTAAATTTAAATGACGTAAAGTTAGATAT 352
QY 121 TATTACACAAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATG 180
DB 351 TATTACACAAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATG 292
QY 181 TTAGGAATAGCTATGTTGATACACAGGACAACCTTCTGTTGTTGTTGTTGTTGTT 240
DB 291 TTAGGAATAGCTATGTTGATACACAGGACAACCTTCTGTTGTTGTTGTTGTTGTT 232
QY 241 GCAAGCCCAACATCAACCTATGATACATATCTGGA 275
DB 231 GCAAGCCCAACATCAACCTATGATACATATGTTGA 197
```

```
ORIGIN
Query Match 35.2%; Score 270.2; DB 6; Length 486;
Best Local Similarity 98.9%; Pred. No. 1.5e-56;
Matches 272; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 ATGTCAGTTGAATCTCAAACTCTAACAATCAGCACAACAACTCAATTACACCAATA 60
DB 16 ATGTCAGTTGAATTTTACAATCTAACAATCAGCACAACAACTCAATTACACCAATA 75
QY 61 ATCAAAATTACTAACACATCTGACAGTGATTTAAATTTAAATGACGTAAAGTTAGATAT 120
DB 76 ATCAAAATTACTAACACATCTGACAGTGATTTAAATTTAAATGACGTAAAGTTAGATAT 135
QY 121 TATTACACAAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATG 180
DB 136 TATTACACAAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATG 195
QY 181 TTAGGAATAGCTATGTTGATACACAGGACAACCTTCTGTTGTTGTTGTTGTTGTT 240
DB 196 TTAGGAATAGCTATGTTGATACACAGGACAACCTTCTGTTGTTGTTGTTGTTGTT 255
QY 241 GCAAGCCCAACATCAACCTATGATACATATCTGGA 275
DB 256 GCAAGCCCAACATCAACCTATGATACATATGTTGA 290
RESULT 4
AR001083/C
LOCUS AR001083 486 bp DNA linear PAT 04-DEC-1998
DEFINITION Sequence 3 from patent US 5738984.
ACCESSION AR001083
VERSION AR001083.1 GI:3963150
KEYWORDS
SOURCE
ORGANISM
REFERENCE 1 (bases 1 to 486)
AUTHORS Shoseyov, O.
TITLE Kits and methods of detection using cellulose binding domain fusion proteins
JOURNAL Patent: US 5738984-A 3 14-APR-1998;
FEATURES
Location/Qualifiers
source 124 a 79 c 89 g 194 t
BASE COUNT 124 a 79 c 89 g 194 t
ORIGIN
Query Match 35.2%; Score 270.2; DB 6; Length 486;
Best Local Similarity 98.9%; Pred. No. 1.5e-56;
Matches 272; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 ATGTCAGTTGAATCTCAAACTCTAACAATCAGCACAACAACTCAATTACACCAATA 60
DB 471 ATGTCAGTTGAATTTTACAATCTAACAATCAGCACAACAACTCAATTACACCAATA 412
QY 61 ATCAAAATTACTAACACATCTGACAGTGATTTAAATTTAAATGACGTAAAGTTAGATAT 120
DB 411 ATCAAAATTACTAACACATCTGACAGTGATTTAAATTTAAATGACGTAAAGTTAGATAT 352
QY 121 TATTACACAAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATG 180
DB 351 TATTACACAAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATG 292
QY 181 TTAGGAATAGCTATGTTGATACACAGGACAACCTTCTGTTGTTGTTGTTGTTGTT 240
DB 291 TTAGGAATAGCTATGTTGATACACAGGACAACCTTCTGTTGTTGTTGTTGTTGTT 232
QY 241 GCAAGCCCAACATCAACCTATGATACATATCTGGA 275
DB 231 GCAAGCCCAACATCAACCTATGATACATATGTTGA 197
```

RESULT 5
AR027448
LOCUS AR027448 486 bp DNA linear PAT 29-SEP-1999
DEFINITION Sequence 1 from patent US 5856201.
ACCESSION AR027448
VERSION AR027448.1 GI:5938268
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE
1 (bases 1 to 486)
Shoseyov,O., Shpiegl,I., Goldstein,M.A. and Doi,R.H.
TITLE Methods of detection using a cellulose binding domain fusion product
JOURNAL Patent: US 5856201-A 1 05-JAN-1999;
FEATURES Location/Qualifiers
source 1..486
/organism="unknown"
BASE COUNT 194 a 89 c 79 g 124 t
ORIGIN
Query Match 35.2%; Score 270.2; DB 6; Length 486;
Best Local Similarity 98.9%; Pred. No. 1.5e-56;
Matches 272; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 ATGTCAGTTGAATTTACAACTCTAACAAATCAGCACAACAACTCAATTACACCAATA 60
DB 16 ATGTCAGTTGAATTTACAACTCTAACAAATCAGCACAACAACTCAATTACACCAATA 75
QY 61 ATCAAAATTAACACATCTGACAGTGAATTTAAATTTAAATGAGCTAAAGTTAGATAT 120
DB 76 ATCAAAATTAACACATCTGACAGTGAATTTAAATTTAAATGAGCTAAAGTTAGATAT 135
QY 121 TATTACACAAGTGATGGTACACAAGGACAAACTTTCTGGTGTGACCATGCTGGTGCATTA 180
DB 136 TATTACACAAGTGATGGTACACAAGGACAAACTTTCTGGTGTGACCATGCTGGTGCATTA 195
QY 181 TTAGGAATAGCTATGTTGATAAACAAGTACAGCAAACTTCGTTAAAGAAACA 240
DB 196 TTAGGAATAGCTATGTTGATAAACAAGTACAGCAAACTTCGTTAAAGAAACA 255
QY 241 GCAAGCCCAACATCAACCTATGATACATATCTGGA 275
DB 256 GCAAGCCCAACATCAACCTATGATACATATGTTGA 290
RESULT 6
AR027449/c
LOCUS AR027449 486 bp DNA linear PAT 29-SEP-1999
DEFINITION Sequence 3 from patent US 5856201.
ACCESSION AR027449
VERSION AR027449.1 GI:5938269
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE
1 (bases 1 to 486)
Shoseyov,O., Shpiegl,I., Goldstein,M.A. and Doi,R.H.
TITLE Methods of detection using a cellulose binding domain fusion product
JOURNAL Patent: US 5856201-A 3 05-JAN-1999;
FEATURES Location/Qualifiers
source 1..486
/organism="unknown"
BASE COUNT 124 a 79 c 89 g 194 t
ORIGIN
Query Match 35.2%; Score 270.2; DB 6; Length 486;
Best Local Similarity 98.9%; Pred. No. 1.5e-56;
Matches 272; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 ATGTCAGTTGAATTTACAACTCTAACAAATCAGCACAACAACTCAATTACACCAATA 60
DB 16 ATGTCAGTTGAATTTACAACTCTAACAAATCAGCACAACAACTCAATTACACCAATA 75
QY 61 ATCAAAATTAACACATCTGACAGTGAATTTAAATTTAAATGAGCTAAAGTTAGATAT 120
DB 76 ATCAAAATTAACACATCTGACAGTGAATTTAAATTTAAATGAGCTAAAGTTAGATAT 135
QY 121 TATTACACAAGTGATGGTACACAAGGACAAACTTTCTGGTGTGACCATGCTGGTGCATTA 180
DB 136 TATTACACAAGTGATGGTACACAAGGACAAACTTTCTGGTGTGACCATGCTGGTGCATTA 195
QY 181 TTAGGAATAGCTATGTTGATAAACAAGTACAGCAAACTTCGTTAAAGAAACA 240
DB 196 TTAGGAATAGCTATGTTGATAAACAAGTACAGCAAACTTCGTTAAAGAAACA 255
QY 241 GCAAGCCCAACATCAACCTATGATACATATCTGGA 275
DB 256 GCAAGCCCAACATCAACCTATGATACATATGTTGA 290
RESULT 7
AR058257
LOCUS AR058257 486 bp DNA linear PAT 29-SEP-1999
DEFINITION Sequence 1 from patent US 5837814.
ACCESSION AR058257
VERSION AR058257.1 GI:5983834
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE
1 (bases 1 to 486)
Shoseyov,O., Shpiegl,I., Goldstein,M. and Doi,R.
TITLE Cellulose binding domain proteins
JOURNAL Patent: US 5837814-A 1 17-NOV-1998;
FEATURES Location/Qualifiers
source 1..486
/organism="unknown"
BASE COUNT 194 a 89 c 79 g 124 t
ORIGIN
Query Match 35.2%; Score 270.2; DB 6; Length 486;
Best Local Similarity 98.9%; Pred. No. 1.5e-56;
Matches 272; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 ATGTCAGTTGAATTTACAACTCTAACAAATCAGCACAACAACTCAATTACACCAATA 60
DB 16 ATGTCAGTTGAATTTACAACTCTAACAAATCAGCACAACAACTCAATTACACCAATA 75
QY 61 ATCAAAATTAACACATCTGACAGTGAATTTAAATTTAAATGAGCTAAAGTTAGATAT 120
DB 76 ATCAAAATTAACACATCTGACAGTGAATTTAAATTTAAATGAGCTAAAGTTAGATAT 135
QY 121 TATTACACAAGTGATGGTACACAAGGACAAACTTTCTGGTGTGACCATGCTGGTGCATTA 180
DB 136 TATTACACAAGTGATGGTACACAAGGACAAACTTTCTGGTGTGACCATGCTGGTGCATTA 195
QY 181 TTAGGAATAGCTATGTTGATAAACAAGTACAGCAAACTTCGTTAAAGAAACA 240
DB 196 TTAGGAATAGCTATGTTGATAAACAAGTACAGCAAACTTCGTTAAAGAAACA 255
QY 241 GCAAGCCCAACATCAACCTATGATACATATCTGGA 275
DB 256 GCAAGCCCAACATCAACCTATGATACATATGTTGA 290
RESULT 8
AR058258/c
LOCUS AR058258 486 bp DNA linear PAT 29-SEP-1999
DEFINITION Sequence 3 from patent US 5837814.
ACCESSION AR058258
VERSION AR058258.1 GI:5983835
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.

Db 471 ATGTCAGTTGAATTTACAACTCTAACAAATCAGCACAACAACTCAATTACACCAATA 412
QY 61 ATCAAAATTAACACATCTGACAGTGAATTTAAATTTAAATGAGCTAAAGTTAGATAT 120
Db 411 ATCAAAATTAACACATCTGACAGTGAATTTAAATTTAAATGAGCTAAAGTTAGATAT 352
QY 121 TATTACACAAGTGATGGTACACAAGGACAAACTTTCTGGTGTGACCATGCTGGTGCATTA 180
Db 351 TATTACACAAGTGATGGTACACAAGGACAAACTTTCTGGTGTGACCATGCTGGTGCATTA 292
QY 181 TTAGGAATAGCTATGTTGATAAACAAGTACAGCAAACTTCGTTAAAGAAACA 240
Db 291 TTAGGAATAGCTATGTTGATAAACAAGTACAGCAAACTTCGTTAAAGAAACA 232
QY 241 GCAAGCCCAACATCAACCTATGATACATATCTGGA 275
Db 231 GCAAGCCCAACATCAACCTATGATACATATGTTGA 197

Unclassified.
1 (bases 1 to 486)
Shoseyov, O., Shpiegl, I., Goldstein, M. A. and Doi, R. H.
Cellulose binding domain proteins
JOURNAL Patent: US 5837814-A 3 17-NOV-1998;
FEATURES
Source Location/Qualifiers
1..486
/organism="unknown"
BASE COUNT 124 a 79 c 89 g 194 t
ORIGIN

Query Match 35.2%; Score 270.2; DB 6; Length 486;
Best Local Similarity 98.9%; Pred. No. 1.5e-56;
Matches 272; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ATGTCAGTTGAATTTACAACTCTAACAAATCAGCACAACAACTCAATTTACACCAATA 60
|||||
Db 471 ATGTCAGTTGAATTTACAACTCTAACAAATCAGCACAACAACTCAATTTACACCAATA 412
|||||

QY 61 ATCAAAATTACTAACACATCTGCAGTGATTTAAATTTAAATGACGTAAAGTTAGATAT 120
|||||

Db 411 ATCAAAATTACTAACACATCTGCAGTGATTTAAATTTAAATGACGTAAAGTTAGATAT 352
|||||

QY 121 TATTACACAAGTGATGTTACACAAGGACAACTTTCTGGTGTGACCATGCTGGTGCAATTA 180
|||||

Db 351 TATTACACAAGTGATGTTACACAAGGACAACTTTCTGGTGTGACCATGCTGGTGCAATTA 292
|||||

QY 181 TTAGGAATAGCTATGTTGATAACACTAGCAAAAGTGACAGCAAACTCAATTTACACCAATA 240
|||||

Db 291 TTAGGAATAGCTATGTTGATAACACTAGCAAAAGTGACAGCAAACTCAATTTACACCAATA 232
|||||

QY 241 GCAAGCCCAACATCAACCTATGATATATCTGGA 275
|||||

Db 231 GCAAGCCCAACATCAACCTATGATATATGTTGA 197
|||||

RESULT 9
I18514
LOCUS I18514
DEFINITION Sequence 1 from patent US 5496934.
ACCESSION I18514
VERSION I18514.1 GI:1598869
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 486)
AUTHORS Shoseyov, O., Shpiegl, I., Goldstein, M. A. and Doi, R. H.
TITLE Nucleic acids encoding a cellulose binding domain
JOURNAL Patent: US 5496934-A 1 05-MAR-1996;
FEATURES
Source Location/Qualifiers
1..486
/organism="unknown"
BASE COUNT 194 a 89 c 79 g 124 t
ORIGIN

Query Match 35.2%; Score 270.2; DB 6; Length 486;
Best Local Similarity 98.9%; Pred. No. 1.5e-56;
Matches 272; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ATGTCAGTTGAATTTACAACTCTAACAAATCAGCACAACAACTCAATTTACACCAATA 60
|||||

Db 16 ATGTCAGTTGAATTTACAACTCTAACAAATCAGCACAACAACTCAATTTACACCAATA 75
|||||

QY 61 ATCAAAATTACTAACACATCTGCAGTGATTTAAATTTAAATGACGTAAAGTTAGATAT 120
|||||

Db 76 ATCAAAATTACTAACACATCTGCAGTGATTTAAATTTAAATGACGTAAAGTTAGATAT 135
|||||

QY 121 TATTACACAAGTGATGTTACACAAGGACAACTTTCTGGTGTGACCATGCTGGTGCAATTA 180
|||||

Db 136 TATTACACAAGTGATGTTACACAAGGACAACTTTCTGGTGTGACCATGCTGGTGCAATTA 195
|||||

QY 181 TTAGGAATAGCTATGTTGATAACACTAGCAAAAGTGACAGCAAACTCAATTTACACCAATA 240
|||||

|||||
Db 196 TTAGGAATAGCTATGTTGATAACACTAGCAAAAGTGACAGCAAACTTCGTTAAAGAAACA 255
|||||

QY 241 GCAAGCCCAACATCAACCTATGATATATCTGGA 275
|||||

Db 256 GCAAGCCCAACATCAACCTATGATATATGTTGA 290
|||||

RESULT 10
I18515/c
LOCUS I18515
DEFINITION Sequence 3 from patent US 5496934.
ACCESSION I18515
VERSION I18515.1 GI:1598870
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 486)
AUTHORS Shoseyov, O., Shpiegl, I., Goldstein, M. A. and Doi, R. H.
TITLE Nucleic acids encoding a cellulose binding domain
JOURNAL Patent: US 5496934-A 3 05-MAR-1996;
FEATURES
Source Location/Qualifiers
1..486
/organism="unknown"
BASE COUNT 124 a 79 c 89 g 194 t
ORIGIN

Query Match 35.2%; Score 270.2; DB 6; Length 486;
Best Local Similarity 98.9%; Pred. No. 1.5e-56;
Matches 272; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ATGTCAGTTGAATTTACAACTCTAACAAATCAGCACAACAACTCAATTTACACCAATA 60
|||||

Db 471 ATGTCAGTTGAATTTACAACTCTAACAAATCAGCACAACAACTCAATTTACACCAATA 412
|||||

QY 61 ATCAAAATTACTAACACATCTGCAGTGATTTAAATTTAAATGACGTAAAGTTAGATAT 120
|||||

Db 411 ATCAAAATTACTAACACATCTGCAGTGATTTAAATTTAAATGACGTAAAGTTAGATAT 352
|||||

QY 121 TATTACACAAGTGATGTTACACAAGGACAACTTTCTGGTGTGACCATGCTGGTGCAATTA 180
|||||

Db 351 TATTACACAAGTGATGTTACACAAGGACAACTTTCTGGTGTGACCATGCTGGTGCAATTA 292
|||||

QY 181 TTAGGAATAGCTATGTTGATAACACTAGCAAAAGTGACAGCAAACTTCGTTAAAGAAACA 240
|||||

Db 291 TTAGGAATAGCTATGTTGATAACACTAGCAAAAGTGACAGCAAACTTCGTTAAAGAAACA 232
|||||

QY 241 GCAAGCCCAACATCAACCTATGATATATCTGGA 275
|||||

Db 231 GCAAGCCCAACATCAACCTATGATATATGTTGA 197
|||||

RESULT 11
I18515/c
LOCUS I18515
DEFINITION Sequence 1 from patent US 5670623.
ACCESSION I18515
VERSION I18515.1 GI:2724592
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 486)
AUTHORS Shoseyov, O., Shpiegl, I., Goldstein, M. A. and Doi, R. H.
TITLE Methods of use of cellulose binding domain proteins
JOURNAL Patent: US 5670623-A 1 23-SEP-1997;
FEATURES
Source Location/Qualifiers
1..486
/organism="unknown"
BASE COUNT 194 a 89 c 79 g 124 t
ORIGIN

Query Match 35.2%; Score 270.2; DB 6; Length 486;
Best Local Similarity 98.9%; Pred. No. 1.5e-56;
Matches 272; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ATGTCAGTTGAATTTACAACTCTAACAAATCAGCACAACAACTCAATTACACCAATA 60
|||||
Db 16 ATGTCAGTTGAATTTACAACTCTAACAAATCAGCACAACAACTCAATTACACCAATA 75
|||||

QY 61 ATCAAAATTAACACATCTGACAGTGATTTAAATTTAAATGAGCTAAAGTTAGATAT 120
|||||
Db 76 ATCAAAATTAACACATCTGACAGTGATTTAAATTTAAATGAGCTAAAGTTAGATAT 135
|||||

QY 121 TATTACAAAGTGATGGTACAAAGCAAACTTTCTGGTGTGACCATCTGGTGCATTA 180
|||||
Db 136 TATTACAAAGTGATGGTACAAAGCAAACTTTCTGGTGTGACCATCTGGTGCATTA 195
|||||

QY 181 TTAGGAAATAGCTATGTTGATAACACTAGCAAGTGACAGCAAACTTCGTTAAAGAAACA 240
|||||
Db 196 TTAGGAAATAGCTATGTTGATAACACTAGCAAGTGACAGCAAACTTCGTTAAAGAAACA 255
|||||

QY 241 GCAAGCCCAACATCAACCTATGATACATATCTGGA 275
|||||
Db 256 GCAAGCCCAACATCAACCTATGATACATATGTTGA 290
|||||

RESULT 12
166615/c
LOCUS 166615 486 bp DNA linear PAT 29-DEC-1997
DEFINITION Sequence 3 from patent US 5670623.
ACCESSION 166615
VERSION 166615.1 GI:2724593
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 486)
AUTHORS Shoseyov,O., Shpiegel,I., Goldstein,M.A. and Doi,R.H.
TITLE Methods of use of cellulose binding domain proteins
JOURNAL Patent: US 5670623-A 3 23-SEP-1997;
FEATURES Location/Qualifiers
source 1..486
BASE COUNT 124 a 79 c 89 g 194 t
ORIGIN

Query Match 35.2%; Score 270.2; DB 6; Length 486;
Best Local Similarity 98.9%; Pred. No. 1.5e-56;
Matches 272; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ATGTCAGTTGAATTTACAACTCTAACAAATCAGCACAACAACTCAATTACACCAATA 60
|||||
Db 471 ATGTCAGTTGAATTTACAACTCTAACAAATCAGCACAACAACTCAATTACACCAATA 412
|||||

QY 61 ATCAAAATTAACACATCTGACAGTGATTTAAATTTAAATGAGCTAAAGTTAGATAT 120
|||||
Db 411 ATCAAAATTAACACATCTGACAGTGATTTAAATTTAAATGAGCTAAAGTTAGATAT 352
|||||

QY 121 TATTACAAAGTGATGGTACAAAGCAAACTTTCTGGTGTGACCATCTGGTGCATTA 180
|||||
Db 351 TATTACAAAGTGATGGTACAAAGCAAACTTTCTGGTGTGACCATCTGGTGCATTA 292
|||||

QY 181 TTAGGAAATAGCTATGTTGATAACACTAGCAAGTGACAGCAAACTTCGTTAAAGAAACA 240
|||||
Db 291 TTAGGAAATAGCTATGTTGATAACACTAGCAAGTGACAGCAAACTTCGTTAAAGAAACA 232
|||||

QY 241 GCAAGCCCAACATCAACCTATGATACATATCTGGA 275
|||||
Db 231 GCAAGCCCAACATCAACCTATGATACATATGTTGA 197
|||||

RESULT 13
168789
LOCUS 168789 486 bp DNA linear PAT 10-AUG-1998

DEFINITION Sequence 1 from patent US 5719044.
ACCESSION 188789
VERSION 188789.1 GI:3408729
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 486)
AUTHORS Shoseyov,O., Shpiegel,I., Goldstein,M.A. and Doi,R.H.
TITLE Cellulose binding domain fusion proteins
JOURNAL Patent: US 5719044-A 1 17-FEB-1998;
FEATURES Location/Qualifiers
source 1..486
BASE COUNT 194 a 89 c 79 g 124 t
ORIGIN

Query Match 35.2%; Score 270.2; DB 6; Length 486;
Best Local Similarity 98.9%; Pred. No. 1.5e-56;
Matches 272; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ATGTCAGTTGAATTTACAACTCTAACAAATCAGCACAACAACTCAATTACACCAATA 60
|||||
Db 16 ATGTCAGTTGAATTTACAACTCTAACAAATCAGCACAACAACTCAATTACACCAATA 75
|||||

QY 61 ATCAAAATTAACACATCTGACAGTGATTTAAATTTAAATGAGCTAAAGTTAGATAT 120
|||||
Db 76 ATCAAAATTAACACATCTGACAGTGATTTAAATTTAAATGAGCTAAAGTTAGATAT 135
|||||

QY 121 TATTACAAAGTGATGGTACAAAGCAAACTTTCTGGTGTGACCATCTGGTGCATTA 180
|||||
Db 136 TATTACAAAGTGATGGTACAAAGCAAACTTTCTGGTGTGACCATCTGGTGCATTA 195
|||||

QY 181 TTAGGAAATAGCTATGTTGATAACACTAGCAAGTGACAGCAAACTTCGTTAAAGAAACA 240
|||||
Db 196 TTAGGAAATAGCTATGTTGATAACACTAGCAAGTGACAGCAAACTTCGTTAAAGAAACA 255
|||||

QY 241 GCAAGCCCAACATCAACCTATGATACATATCTGGA 275
|||||
Db 256 GCAAGCCCAACATCAACCTATGATACATATGTTGA 290
|||||

RESULT 14
188790/c
LOCUS 188790 486 bp DNA linear PAT 10-AUG-1998
DEFINITION Sequence 3 from patent US 5719044.
ACCESSION 188790
VERSION 188790.1 GI:3408730
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 486)
AUTHORS Shoseyov,O., Shpiegel,I., Goldstein,M.A. and Doi,R.H.
TITLE Cellulose binding domain fusion proteins
JOURNAL Patent: US 5719044-A 3 17-FEB-1998;
FEATURES Location/Qualifiers
source 1..486
BASE COUNT 124 a 79 c 89 g 194 t
ORIGIN

Query Match 35.2%; Score 270.2; DB 6; Length 486;
Best Local Similarity 98.9%; Pred. No. 1.5e-56;
Matches 272; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ATGTCAGTTGAATTTACAACTCTAACAAATCAGCACAACAACTCAATTACACCAATA 60
|||||
Db 471 ATGTCAGTTGAATTTACAACTCTAACAAATCAGCACAACAACTCAATTACACCAATA 412
|||||

QY 61 ATCAAAATTAACACATCTGACAGTGATTTAAATTTAAATGAGCTAAAGTTAGATAT 120
|||||
Db 471 ATGTCAGTTGAATTTACAACTCTAACAAATCAGCACAACAACTCAATTACACCAATA 412
|||||

QY 61 ATCAAAATTAACACATCTGACAGTGATTTAAATTTAAATGAGCTAAAGTTAGATAT 120
|||||
Db 411 ATCAAAATTAACACATCTGACAGTGATTTAAATTTAAATGAGCTAAAGTTAGATAT 352
|||||

QY 121 TATTACACAAGTGTGTTACACAGGACAAACTTTCTGGTGTGACCATGCTGGTGCAATTA 180
|||||
Db 351 TATTACACAAGTGTGTTACACAGGACAAACTTTCTGGTGTGACCATGCTGGTGCAATTA 292
|||||
QY 181 TTAGGAAATAGCTATGTTGATAACACTAGCAAAAGTGACAGCAAACTTCGTTAAAGAAACA 240
|||||
Db 291 TTAGGAAATAGCTATGTTGATAACACTAGCAAAAGTGACAGCAAACTTCGTTAAAGAAACA 232
|||||
QY 241 GCAAGCCCAACATCAACCTATGATACATATCTGGA 275
|||||
Db 231 GCAAGCCCAACATCAACCTATGATACATATCTGGA 197
|||||

RESULT 15

AR096204 LOCUS AR096204 499 bp DNA linear PAT 08-SEP-2000
DEFINITION Sequence 5 from patent US 6005092.

ACCESSION AR096204

VERSION AR096204.1 GI:10024795

KEYWORDS

SOURCE Unknown.

ORGANISM Unknown.

REFERENCE 1 (bases 1 to 499)

AUTHORS Shoseyov, O. and Shani, Z.

TITLE Arabidopsis thaliana endo-1,4-beta-D-glucanase gene and promoter

JOURNAL Patent: US 6005092-A 5 21-DEC-1999;

FEATURES Location/Qualifiers

Source

1..499

/organism="unknown"

BASE COUNT 197 a 93 c 82 g 127 t

ORIGIN

Query Match

35.2%; Score 270.2; DB 6; Length 499;

Best Local Similarity 98.9%; Pred. No. 1.5e-56;

Matches 272; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ATGTCAGTTGAATTTACAACTCTAACAAATCAGCACAAACAACTCAATTACACCAATA 60
|||||Db 21 ATGTCAGTTGAATTTACAACTCTAACAAATCAGCACAAACAACTCAATTACACCAATA 80
|||||QY 61 ATCAAAATTAACACATCTGACAGTGTAAATTTAAATGACGTAAAGTTAGATAT 120
|||||Db 81 ATCAAAATTAACACATCTGACAGTGTAAATTTAAATGACGTAAAGTTAGATAT 140
|||||QY 121 TATTACACAAGTGTGTTACACAGGACAAACTTTCTGGTGTGACCATGCTGGTGCAATTA 180
|||||Db 141 TATTACACAAGTGTGTTACACAGGACAAACTTTCTGGTGTGACCATGCTGGTGCAATTA 200
|||||QY 181 TTAGGAAATAGCTATGTTGATAACACTAGCAAAAGTGACAGCAAACTTCGTTAAAGAAACA 240
|||||Db 201 TTAGGAAATAGCTATGTTGATAACACTAGCAAAAGTGACAGCAAACTTCGTTAAAGAAACA 260
|||||QY 241 GCAAGCCCAACATCAACCTATGATACATATCTGGA 275
|||||Db 261 GCAAGCCCAACATCAACCTATGATACATATCTGGA 295
|||||

Search completed: February 22, 2003, 16:18:12

Job time : 2186.73 secs

THIS PAGE BLANK (USPTO)

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 | 26 | 27 | 28 | 29 | 30 | 31 | 32 | 33 | 34 | 35 | 36 | 37 | 38 | 39 | 40 | 41 | 42 | 43 | 44 | 45 | 46 | 47 | 48 | 49 | 50 | 51 | 52 | 53 | 54 | 55 | 56 | 57 | 58 | 59 | 60 | 61 | 62 | 63 | 64 | 65 | 66 | 67 | 68 | 69 | 70 | 71 | 72 | 73 | 74 | 75 | 76 | 77 | 78 | 79 | 80 | 81 | 82 | 83 | 84 | 85 | 86 | 87 | 88 | 89 | 90 | 91 | 92 | 93 | 94 | 95 | 96 | 97 | 98 | 99 | 100 | 101 | 102 | 103 | 104 | 105 | 106 | 107 | 108 | 109 | 110 | 111 | 112 | 113 | 114 | 115 | 116 | 117 | 118 | 119 | 120 | 121 | 122 | 123 | 124 | 125 | 126 | 127 | 128 | 129 | 130 | 131 | 132 | 133 | 134 | 135 | 136 | 137 | 138 | 139 | 140 | 141 | 142 | 143 | 144 | 145 | 146 | 147 | 148 | 149 | 150 | 151 | 152 | 153 | 154 | 155 | 156 | 157 | 158 | 159 | 160 | 161 | 162 | 163 | 164 | 165 | 166 | 167 | 168 | 169 | 170 | 171 | 172 | 173 | 174 | 175 | 176 | 177 | 178 | 179 | 180 | 181 | 182 | 183 | 184 | 185 | 186 | 187 | 188 | 189 | 190 | 191 | 192 | 193 | 194 | 195 | 196 | 197 | 198 | 199 | 200 | 201 | 202 | 203 | 204 | 205 | 206 | 207 | 208 | 209 | 210 | 211 | 212 | 213 | 214 | 215 | 216 | 217 | 218 | 219 | 220 | 221 | 222 | 223 | 224 | 225 | 226 | 227 | 228 | 229 | 230 | 231 | 232 | 233 | 234 | 235 | 236 | 237 | 238 | 239 | 240 | 241 | 242 | 243 | 244 | 245 | 246 | 247 | 248 | 249 | 250 | 251 | 252 | 253 | 254 | 255 | 256 | 257 | 258 | 259 | 260 | 261 | 262 | 263 | 264 | 265 | 266 | 267 | 268 | 269 | 270 | 271 | 272 | 273 | 274 | 275 | 276 | 277 | 278 | 279 | 280 | 281 | 282 | 283 | 284 | 285 | 286 | 287 | 288 | 289 | 290 | 291 | 292 | 293 | 294 | 295 | 296 | 297 | 298 | 299 | 300 | 301 | 302 | 303 | 304 | 305 | 306 | 307 | 308 | 309 | 310 | 311 | 312 | 313 | 314 | 315 | 316 | 317 | 318 | 319 | 320 | 321 | 322 | 323 | 324 | 325 | 326 | 327 | 328 | 329 | 330 | 331 | 332 | 333 | 334 | 335 | 336 | 337 | 338 | 339 | 340 | 341 | 342 | 343 | 344 | 345 | 346 | 347 | 348 | 349 | 350 | 351 | 352 | 353 | 354 | 355 | 356 | 357 | 358 | 359 | 360 | 361 | 362 | 363 | 364 | 365 | 366 | 367 | 368 | 369 | 370 | 371 | 372 | 373 | 374 | 375 | 376 | 377 | 378 | 379 | 380 | 381 | 382 | 383 | 384 | 385 | 386 | 387 | 388 | 389 | 390 | 391 | 392 | 393 | 394 | 395 | 396 | 397 | 398 | 399 | 400 | 401 | 402 | 403 | 404 | 405 | 406 | 407 | 408 | 409 | 410 | 411 | 412 | 413 | 414 | 415 | 416 | 417 | 418 | 419 | 420 | 421 | 422 | 423 | 424 | 425 | 426 | 427 | 428 | 429 | 430 | 431 | 432 | 433 | 434 | 435 | 436 | 437 | 438 | 439 | 440 | 441 | 442 | 443 | 444 | 445 | 446 | 447 | 448 | 449 | 450 | 451 | 452 | 453 | 454 | 455 | 456 | 457 | 458 | 459 | 460 | 461 | 462 | 463 | 464 | 465 | |
|--|---|---|---|---|---|---|---|---|---|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|--|
|--|---|---|---|---|---|---|---|---|---|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|--|

PN CA2281913
yy

| PN | CA2 |
|-----|-----|
| 1 | 1 |
| 2 | 2 |
| 3 | 3 |
| 4 | 4 |
| 5 | 5 |
| 6 | 6 |
| 7 | 7 |
| 8 | 8 |
| 9 | 9 |
| 10 | 10 |
| 11 | 11 |
| 12 | 12 |
| 13 | 13 |
| 14 | 14 |
| 15 | 15 |
| 16 | 16 |
| 17 | 17 |
| 18 | 18 |
| 19 | 19 |
| 20 | 20 |
| 21 | 21 |
| 22 | 22 |
| 23 | 23 |
| 24 | 24 |
| 25 | 25 |
| 26 | 26 |
| 27 | 27 |
| 28 | 28 |
| 29 | 29 |
| 30 | 30 |
| 31 | 31 |
| 32 | 32 |
| 33 | 33 |
| 34 | 34 |
| 35 | 35 |
| 36 | 36 |
| 37 | 37 |
| 38 | 38 |
| 39 | 39 |
| 40 | 40 |
| 41 | 41 |
| 42 | 42 |
| 43 | 43 |
| 44 | 44 |
| 45 | 45 |
| 46 | 46 |
| 47 | 47 |
| 48 | 48 |
| 49 | 49 |
| 50 | 50 |
| 51 | 51 |
| 52 | 52 |
| 53 | 53 |
| 54 | 54 |
| 55 | 55 |
| 56 | 56 |
| 57 | 57 |
| 58 | 58 |
| 59 | 59 |
| 60 | 60 |
| 61 | 61 |
| 62 | 62 |
| 63 | 63 |
| 64 | 64 |
| 65 | 65 |
| 66 | 66 |
| 67 | 67 |
| 68 | 68 |
| 69 | 69 |
| 70 | 70 |
| 71 | 71 |
| 72 | 72 |
| 73 | 73 |
| 74 | 74 |
| 75 | 75 |
| 76 | 76 |
| 77 | 77 |
| 78 | 78 |
| 79 | 79 |
| 80 | 80 |
| 81 | 81 |
| 82 | 82 |
| 83 | 83 |
| 84 | 84 |
| 85 | 85 |
| 86 | 86 |
| 87 | 87 |
| 88 | 88 |
| 89 | 89 |
| 90 | 90 |
| 91 | 91 |
| 92 | 92 |
| 93 | 93 |
| 94 | 94 |
| 95 | 95 |
| 96 | 96 |
| 97 | 97 |
| 98 | 98 |
| 99 | 99 |
| 100 | 100 |

XX Claim 16; Fig 5; 55pp; English.

XX The invention describes a method of protecting a poikilothermic fish

CC against infection by the bacterial pathogen *Piscirickettsia salmonis*

CC comprising administering either intraperitoneally, by immersion or

CC orally, an immunogenic amount of principal antigen, the Ospa (outer

CC surface lipoprotein), its variants, non-lipidated form or antigenic

CC peptides derived or synthesized with or without an adjuvant. The new

CC method is used to provide an outer surface lipoprotein (Ospa) of

CC bacterial strain *Piscirickettsia salmonis* as a vaccine to induce immunity

CC in fin-fish against *Rickettsial* septicemia and other related

CC *Rickettsial* diseases caused by either a virus, bacteria or parasite.

CC This sequence encodes the *Escherichia coli* codon optimised outer surface

CC lipoprotein Ospa (17ex) with an N-terminal fusion used in the creation of

CC the vaccine described in the invention.

XX

SQ Sequence 768 BP; 220 A; 183 C; 190 G; 175 T; 0 other;

Query Match 100.0%; Score 768; DB 24; Length 768;

Best Local Similarity 100.0%; Pred. No. 1.4e-214;

Matches 768; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGTCAGTTGAATCTCAACTCTAACAATCAGCAACAACAACTCAATTACACCAATA 60

DB 1 ATGTCAGTTGAATCTCAACTCTAACAATCAGCAACAACAACTCAATTACACCAATA 60

QY 61 ATCAAAATTAATAACATCTGACAGTGAATTAATTAATGACGTAAAGTTAGATAT 120

DB 61 ATCAAAATTAATAACATCTGACAGTGAATTAATTAATGACGTAAAGTTAGATAT 120

QY 121 TATTACACAGTGTGTACACAGGACAACTTCTGGTGTGACCATGCTGGTGCAATTA 180

DB 121 TATTACACAGTGTGTACACAGGACAACTTCTGGTGTGACCATGCTGGTGCAATTA 180

QY 181 TTAGGAATAGCTATGTTGTAACACTAGCAAGTGACGACAACTTCGTTAAAGAAACA 240

DB 181 TTAGGAATAGCTATGTTGTAACACTAGCAAGTGACGACAACTTCGTTAAAGAAACA 240

QY 241 GCAAGCCCAACATCAAGCTATGATACATATCTGGATCCGTCTCATATGCGTGGTTCG 300

DB 241 GCAAGCCCAACATCAAGCTATGATACATATCTGGATCCGTCTCATATGCGTGGTTCG 300

QY 301 CAGGCGAGCTCTGATCATATCTCTGTTTCTGTTGGTGGCCGACAACTTCAGC 360

DB 301 CAGGCGAGCTCTGATCATATCTCTGTTTCTGTTGGTGGCCGACAACTTCAGC 360

QY 361 CGCCAGGAAGTTGGCGGGCCACCGGTGCGGTGTTGGCGGTGTTGGCGGACGCTGTC 420

DB 361 CGCCAGGAAGTTGGCGGGCCACCGGTGCGGTGTTGGCGGTGTTGGCGGACGCTGTC 420

QY 421 GGTAAAGCTCTGTTGTTGTCATGATGATGATGATGATGATGATGATGATGATGATGAT 480

DB 421 GGTAAAGCTCTGTTGTTGTCATGATGATGATGATGATGATGATGATGATGATGATGAT 480

QY 481 GGCTCTAAATCGGTGACAGATGACGACGACGATGATGATGATGATGATGATGATGATGAT 540

DB 481 GGCTCTAAATCGGTGACAGATGACGACGACGATGATGATGATGATGATGATGATGATGAT 540

QY 541 GAAAGAGTGAAGCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 600

DB 541 GAAAGAGTGAAGCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 600

QY 601 TCTGTGAACCGGTTGCGACCTACCGAGTGTACAAACAAACAGGAACCGGTCAGCAGTAC 660

DB 601 TCTGTGAACCGGTTGCGACCTACCGAGTGTACAAACAAACAGGAACCGGTCAGCAGTAC 660

QY 661 TGCCGCAATTCAGAGAAAGCCATGATGCGAGGTGAGAAACAGGAATTCAGCGGAC 720

DB 661 TGCCGCAATTCAGAGAAAGCCATGATGCGAGGTGAGAAACAGGAATTCAGCGGAC 720

QY 721 GCGTGCCTCAGCGGGATGGCGGTGCGAGGTGATGATGATGATGATGATGATGATGATGAT 768

DB 721 GCGTGCCTCAGCGGGATGGCGGTGCGAGGTGATGATGATGATGATGATGATGATGATGAT 768

Db 721 GCGTGCCTCAGCGGGATGGCGGTGCGAGGTGATGATGATGATGATGATGATGATGAT 768

RESULT 3

AAF86247

ID AAF86247 standard; DNA; 483 BP.

XX AAF86247;

AC AAF86247;

XX 11-JUL-2001 (first entry)

XX DNA sequence of *E. coli* optimised ospA gene 17E2.

DE Poikilothermic fish; *Piscirickettsia salmonis*; rickettsial pathogen;

XX vaccine; Ospa; salmonid rickettsial septicemia; rickettsial disease;

KW SRS; 17E2; ds.

KW

XX *Piscirickettsia salmonis*.

OS Synthetic.

XX

XX Key Location/Qualifiers

PH 1..483

FT CDS

FT /*tag= a

FT /partial

FT /product= "Ospa"

FT /note= "Genus specific 17kDa antigen, the sequence does not include a stop codon"

FT

XX CA2281913-A1.

XX

XX 17-MAR-2001.

XX

XX 17-SEP-1999; 99CA-2281913.

XX

XX 17-SEP-1999; 99CA-2281913.

XX

XX (KAYW/) KAY W W.

PA (BURI/) BURIAN J.

PA (KUYZ/) KUYZ M A.

XX

XX Kay WW, Burian J, Kuzyk MA;

XX

XX WPI; 2001-316844/34.

DR P-PSDB; AAB81127.

DR

XX

XX Method for protecting poikilothermic fish against salmonid rickettsial

PT septicemia and other rickettsial diseases comprises administering a

PT vaccine containing the Ospa protein of *Piscirickettsia salmonis*

XX

PS Example 3; Fig 4C; 35pp; English.

XX

CC This invention relates to a method for the protection against infection

CC of a poikilothermic fish by the bacterial pathogen, *Piscirickettsia*

CC salmonis. The method comprises administering an immunogenic amount of a

CC *P. salmonis* specific antigen termed Ospa, or an immunogenic fragment of

CC Ospa in the form of a vaccine. The method is used for protecting animals,

CC particularly poikilothermic fish, against the bacterial pathogen

CC *P. salmonis*. The method is also useful for protecting against salmonid

CC rickettsial septicemia (SRS) and other rickettsial diseases. The present

CC sequence represents DNA which has been optimised for expression in

CC *Escherichia coli* to encode the *P. salmonis* Ospa protein. An Ospa protein

CC with an N-terminal fusion partner is used in a vaccine to create an

CC anti-Ospa antibody response.

XX

SQ Sequence 483 BP; 108 A; 127 C; 148 G; 100 T; 0 other;

Query Match 62.9%; Score 483; DB 22; Length 483;

Best Local Similarity 100.0%; Pred. No. 3.4e-131;

Matches 483; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 286 ATGCGTGGTTCCTCGCAGGCGAGCTCTCTGATCATATCTCTGTTTCTCTGTTTCTCTGTTTCT 345

DB 1 ATGCGTGGTTCCTCGCAGGCGAGCTCTCTGATCATATCTCTGTTTCTCTGTTTCTCTGTTTCT 60

QY 346 GCCCAGAACTTCAGCCGCGCAGGAAAGTTGGCGGGCCACCCGGTGCCTGCGGTTGTGGCGCGTGT 405
 Db 61 GCCCAGAACTTCAGCCGCGCAGGAAAGTTGGCGGGCCACCCGGTGCCTGCGGTTGTGGCGCGTGT 120
 QY 406 GCGCGCCAGCTGTTCCGTTAAAGCTCTGCTCGTGTGTCATGCCCATCGCGGTCGGGTT 465
 Db 121 GCGCGCCAGCTGTTCCGTTAAAGCTCTGCTCGTGTGTCATGCCCATCGCGGTCGGGTT 180
 QY 466 CTGGCGGCTGCTGATGGCTCTAAATCGGTTCAGAGCATGGACGACGAGGATAAAATCAAA 525
 Db 181 CTGGCGGCTGCTGATGGCTCTAAATCGGTTCAGAGCATGGACGACGAGGATAAAATCAAA 240
 QY 526 CTGAACCACTCTCTGAAAGCTGAAAGCGCGCCAGGTTACTCGTTGGCGTAAATCCGGAC 585
 Db 241 CTGAACCACTCTCTGAAAGCTGAAAGCGCGCCAGGTTACTCGTTGGCGTAAATCCGGAC 300
 QY 586 ACCGGTAAACAGCTACTCTGTTGAACCGGTTCCACCTACCGGTTTACACAAACAGGAA 645
 Db 301 ACCGGTAAACAGCTACTCTGTTGAACCGGTTCCACCTACCGGTTTACACAAACAGGAA 360
 QY 646 CGCGCTCAGCAGTACTGCGCGGAAATTCAGCAGAAAGCCATGATCGCAGGTCAAGAAACAG 705
 Db 361 CGCGCTCAGCAGTACTGCGCGGAAATTCAGCAGAAAGCCATGATCGCAGGTCAAGAAACAG 420
 QY 706 GAAATCTAGCGCACCGCTGCGCTCAGCGGATGCGCGTGGCAGGTGATTAGCACCGAA 765
 Db 421 GAAATCTAGCGCACCGCTGCGCTCAGCGGATGCGCGTGGCAGGTGATTAGCACCGAA 480
 QY 766 AAA 768
 Db 481 AAA 483

RESULT 4

ABK52402
 ID ABK52402 standard; DNA; 483 BP.
 XX
 AC ABK52402;
 XX
 DT 12-AUG-2002 (first entry)
 XX
 DE DNA encoding Escherichia coli codon optimised OspA, 17e2.
 XX
 KW Outer surface lipoprotein; OspA; antibacterial; immunosuppressive;
 KW vaccine; poikilothermic fish; fin-fish; Rickettsial septicemia;
 KW Rickettsial disease; gene; ds; 17e2.
 XX
 OS Piscirickettsia salmonis.
 OS Synthetic.
 XX
 FH Key Location/Qualifiers
 FT CDS 1..483
 FT /*tag= a
 FT /product= "17e2"
 FT /note= "Escherichia coli codon optimised OspA"
 FT /partial
 FT /note= "No stop codon given"

CA2339327-Al.

15-MAR-2002.

PF 19-MAR-2001; 2001CA-2339327.

PR 15-SEP-2000; 2000US-0677374.

(THOR/) THORNTON J C.

PA (KAYW/) KAY W W.

PA (BURI/) BURIAN J.

PA (KUZY/) KUZYSK M A.

PI Thornton JC, Kay WW, Burian J, Kuzysk MA;

XX

DR WPI; 2002-455221/49.

P-PSDB; AAU97868.

XX

PT Inducing immunity in fin fish to Rickettsial septicemia, comprises
 administration of an outer surface lipoprotein (OspA) of a bacterial
 strain, as a vaccine -

PT

PS Claim 16; Fig 4; 55pp; English.

XX

The invention describes a method of protecting a poikilothermic fish
 against infection by the bacterial pathogen *Piscirickettsia salmonis*
 comprising administering either intraperitoneally, by immersion or
 orally, an immunogenic amount of principal antigen, the OspA (outer
 surface lipoprotein), its variants, non-lipidated form or antigenic
 peptides derived or synthesized with or without an adjuvant. The new
 method is used to provide an outer surface lipoprotein (OspA) of
 bacterial strain *Piscirickettsia salmonis* as a vaccine to induce immunity
 in fin-fish against Rickettsial septicemia and other related
 Rickettsial diseases caused by either a virus, bacteria or parasite.
 CC This sequence encodes the *Escherichia coli* codon optimised outer surface
 lipoprotein OspA (17e2) used in the creation of the vaccine described in
 the invention.

SQ Sequence 483 BP; 108 A; 127 C; 148 G; 100 T; 0 other;

Query Match

Best Local Similarity 62.9%; Score 483; DB 24; Length 483;

Matches 483; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY

286 ATGCGTGGTTGCTGCAGGGCAGCTCTCTGATCATATCTCTGTTTCTGTTGGTGGTTC 345
 Db 1 ATGCGTGGTTGCTGCAGGGCAGCTCTCTGATCATATCTCTGTTTCTGTTGGTGGTTC 60

QY

346 GCCCAGAACTTCAGCGCCAGGAAAGTTGGCGGGCCACCCGGTGCCTGCGGTCGGTGT 405
 Db 61 GCCCAGAACTTCAGCGCCAGGAAAGTTGGCGGGCCACCCGGTGCCTGCGGTCGGTGT 120

QY

406 GCGCGCCAGCTGTTCCGTTAAAGCTCTGCTGCTGTGTCATGCCCATCGCGGTCGGTGT 465
 Db 121 GCGCGCCAGCTGTTCCGTTAAAGCTCTGCTGCTGTGTCATGCCCATCGCGGTCGGTGT 180

QY

466 CTGGCGGCTCTGATGGCTCTAAATCGGTTCAGAGCATGGACGACGAGGATAAAATCAAA 525
 Db 181 CTGGCGGCTCTGATGGCTCTAAATCGGTTCAGAGCATGGACGACGAGGATAAAATCAAA 240

QY

526 CTGAACCACTCTCTGAAAGCTGAAAGCGCGCCAGGTTACTCGTTGGCGTAAATCCGGAC 585
 Db 241 CTGAACCACTCTCTGAAAGCTGAAAGCGCGCCAGGTTACTCGTTGGCGTAAATCCGGAC 300

QY

586 ACCGGTAAACAGCTACTCTGTGGAACCGGTTCCACCTACCGGTTTACACAAACAGGAA 645
 Db 301 ACCGGTAAACAGCTACTCTGTGGAACCGGTTCCACCTACCGGTTTACACAAACAGGAA 360

QY

646 GCGCGTCAGCAGTACTGCGCGGAAATTCAGCAGAAAGCCATGATCGAGGTGAGAAACAG 705
 Db 361 GCGCGTCAGCAGTACTGCGCGGAAATTCAGCAGAAAGCCATGATCGAGGTGAGAAACAG 420

QY

706 GAAATCTAGCGCACCGCTGCGCTCAGCGGATGCGCGTGGCAGGTGATTAGCACCGAA 765
 Db 421 GAAATCTAGCGCACCGCTGCGCTCAGCGGATGCGCGTGGCAGGTGATTAGCACCGAA 480

QY 766 AAA 768

Db 481 AAA 483

RESULT 5

AAF86246

ID AAF86246 standard; DNA; 486 BP.

XX

AC AAF86246;

XX

| | | | |
|--------|---|---|-----|
| Db | 188 | GTGGATTAAATGGTTCTTAAATCGGTCAATCGATGGATCAGCAGGATAAAATAAGCTAA | 24 |
| QY | 530 | ACCAGTCTCTGGAAAAAGTGAAGCCGGCCAGGTTACTCGTTGGCGTAAATCCGACACCG | 589 |
| Db | 248 | ACCAGAGTTTTGGAAAAGGTTAAAGAGCAGGCGCAAGTGCACCGTTGGCGTAAATCCAGATACAG | 307 |
| QY | 590 | GTAACAGCTACTCTGTGGAACCGGTTCCGCACCTACACAGCGTTTACACAAACAGGAACGCC | 649 |
| Db | 308 | GCAATAGTTTATAGTGTGTGAGCCAGTGGCTACTTACCAGCGTTTACAATTAAGCAAGAGCGTC | 367 |
| QY | 650 | GTGACGAGTACTGCGCGGAATTTTCAGCAGAAAGCCATGATCGCAGGTCAGAAAACAGGAAA | 709 |
| Db | 368 | GCCAGCAATATTGTCGAGAAATTCAGCAAAAGGCGGATGATTCAGGGCAGAGCAAGAGA | 427 |
| QY | 710 | TCTACGCCACCGGTGCCCTCAGCCGGATGCCGCTGGCAGGTGATTTAGCACCGAAAAA | 768 |
| Db | 428 | TTTACGGCACTGCATGCCGCAACCGGATGCTCGTTGGCAAGTCATTTCAACAGAAAAA | 486 |
| RESULT | 6 | | |
| ID | ABK52401 | | |
| XX | ABK52401 standard; DNA; 486 BP. | | |
| AC | ABK52401; | | |
| XX | | | |
| DT | 12-AUG-2002 (first entry) | | |
| XX | | | |
| DE | DNA encoding outer surface lipoprotein OspA. | | |
| XX | | | |
| KW | Outer surface lipoprotein; OspA; antibacterial; immunosuppressive; | | |
| KW | vaccine; poikilothermic fish; fin-fish; Rickettsial septicaemia; | | |
| KW | Rickettsial disease; gene; ds. | | |
| XX | | | |
| OS | Piscirickettsia salmonis. | | |
| XX | | | |
| FH | Key Location/Qualifiers | | |
| FT | CDS 1..486 | | |
| FT | /*tag= a | | |
| FT | /product= "OspA" | | |
| FT | /note= "Outer surface lipoprotein" | | |
| FT | /partial= "No stop codon given" | | |
| XX | | | |
| PN | CA2339327-AL. | | |
| XX | | | |
| PD | 15-MAR-2002. | | |
| XX | | | |
| PF | 19-MAR-2001; 2001CA-2339327. | | |
| XX | | | |
| PR | 15-SEP-2000; 2000US-0677374. | | |
| XX | | | |
| PA | (THOR/) THORNTON J C. | | |
| PA | (KAYW/) KAY W W. | | |
| PA | (BURI/) BURIAN J. | | |
| PA | (KUZY/) KUZYSK M A. | | |
| XX | | | |
| PI | Thornton JC, Kay WW, Burian J, Kuzyk MA; | | |
| XX | | | |
| DR | WPI: 2002-455221/49. | | |
| DR | P-PSDB; AAU97867. | | |
| XX | | | |
| PT | Inducing immunity in fin fish to Rickettsial septicaemia, comprises | | |
| PT | administration of an outer surface lipoprotein (OspA) of a bacterial | | |
| PT | strain, as a vaccine - | | |
| XX | | | |
| PS | Claim 19; Fig 2: 55pp; English. | | |
| XX | | | |
| CC | The invention describes a method of protecting a poikilothermic fish | | |
| CC | against infection by the bacterial pathogen <i>Piscirickettsia salmonis</i> | | |
| CC | comprising administering either intraperitoneally, by immersion or | | |
| CC | orally, an immunogenic amount of principal antigen, the OspA (outer | | |
| CC | surface lipoprotein), its variants, non-lipidated form or antigenic | | |
| CC | peptides derived or synthesized with or without an adjuvant. The new | | |
| CC | method is used to provide an outer surface lipoprotein (OspA) of | | |

bacterial strain *Piscirickettsia salmonis* as a vaccine to induce immunity in fin-fish against Rickettsial septicaemia and other related Rickettsial diseases caused by either a virus, bacteria or parasite. This sequence encodes the *Piscirickettsia salmonis* outer surface lipoprotein, OspA, used in the creation of the vaccine described in the CC invention.

xx
Sequence 486 BP: 137 A: 79 C: 144 G: 126 T: 0 other:

| | | | | |
|-----------------------|-----------------|------------------|-----------|-------------|
| Query Match | 38.4% | Score 295; | DB 24; | Length 486; |
| Best Local Similarity | 76.0%; | Pred. No. 4e-76; | | |
| Matches 364; | Conservative 0; | Mismatches 115; | Indels 0; | Gaps 0; |

| | | | |
|----|-----|---|-----|
| QY | 290 | GTGGTGGCTCGACGGGAGCTCTCGATCATATCTCTGTTTCTCGTGGTGGTTCGGGCC | 349 |
| DB | | | |
| DB | 8 | GAGATGTTGCAGGTAGTAGTCTAATATATACAGTGTGTTTTTATTGTCGTGGCC | 67 |
| QY | 350 | AGAACTTCAGCCGCCAGGAAGTTGGCGGGCCACCCTGGTTCGGTTCGTGGCGGTGTTCGCG | 409 |
| DB | | | |
| DB | 68 | AGAACTTTAGTCGTCAGAAGTCGGAGTCGGACTGGGGCTGTTGTTGGCGGTGTTCGTG | 127 |
| QY | 410 | GCCAGCTGTTTCGGTAAAGGCTCTGCTGTGTGTCAGTGGCCATCGGCGTGGCGGTTCGCTGG | 469 |
| DB | | | |
| DB | 128 | GCCAGCTGTTTGGTAAAGGTAGTGGTCAGTTGCAATGGCCATTGGTGGTCTGTTTTGG | 187 |
| QY | 470 | CGGCTCGATTTGGCTTAAAAATCGGTCAGAGCATGGACCGAGGATAAAAATCAAACTGA | 529 |
| DB | | | |
| DB | 188 | GTGGATTAATTTGGTTCTAAAAATCGGCTAAATCGATGGATCAGCAGGATAAAAATAAGCTAA | 247 |
| QY | 530 | ACCAGTCTCTCGAAAAAGTGAAGCCGGCCAGGTTACTTCGTTGGCGTAAATCCGGACACCG | 589 |
| DB | | | |
| DB | 248 | ACCAGAGTTTGGAAAAAGTTAAACAGGGCAAGTGCACAGTTGGCGTAAATCCAGATACAG | 307 |
| QY | 590 | GTAACAGCTACTCTGTGGAAACCGGTTCCGACCTTACCAGCGTTTACAACAAACAGGAACGCC | 649 |
| DB | | | |
| DB | 308 | GCAATAGTTATAGTGTTCGAGCCAGTCGCTACTTACCAGCGTTTACAATAAGCAAGAGCGTC | 367 |
| QY | 650 | GTCAGCAGTACTCGCCGGAATTTACAGAGAAGCCATGATCCGAGGTCAGAAACAGGAAA | 709 |
| DB | | | |
| DB | 368 | GCCAGCAATATTGTCGAGAAATTCAGCAAAAGGGCGATGATTCGAGGGCGAAGCAAGAGA | 427 |
| QY | 710 | TCTACGGCACCGGTGCCCTTCAGCCGGATGCCGCTGGCAGGTGATTAGCACCCGAAAAA | 768 |
| DB | | | |
| DB | 428 | TTTACGGCACTGCAATCGCGCAACCGGATGTCGTTGGCAAGTCATTTCAACAGAAAAA | 486 |

[illegible]

```

PR 01-JUL-2000; 2000GB-0016080.
PR 01-JUL-2000; 2000GB-0016082.
PR 29-JUL-2000; 2000GB-0018599.
XX
XX (AQUA-) AQUA HEALTH EURO LTD.
PA
XX Simard N, Brouwers H, Jones S, Griffiths S, Valenzuela P;
PI Burzio L;
XX
XX WPI: 2001-639050/73.
DR P-PSDB: AAG78025.
XX
XX New nucleic acids encoding an amino acid sequence homologous to the
PT surface antigen present on Piscirickettsia salmonis are useful to
PT protect fish against piscirickettsiosis -
XX
XX Claim 3; Fig 4; 25pp; English.
XX
XX The invention relates to nucleic acid sequences and the encoded protein
CC of a least part of the surface antigen present on Piscirickettsia
CC salmonis for production of a vaccine with antibacterial activity to
CC protect fish against P. salmonis which causes piscirickettsiosis, also
CC known as salmonid rickettsial septicaemia.
XX.
SQ Sequence 489 BP; 139 A; 79 C; 144 G; 127 T; 0 other;

Query Match          38.4%; Score 295; DB 22; Length 489;
Best Local Similarity 76.08; Pred. No. 4e-76;
Matches 364; Conservative 0; Mismatches 115; Indels 0; Gaps 0;

QY 290 GTGGTTCCTGCGCAGGCAGCTCTCATATCTCTGTCTTTCTGCTGGTTCGCCGCC 349
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 8 GAGGATGTTTGCAAGTAGTAGTCTAATTATATACAGTGTGTTTTAGTTGGCTGTGCC 67

QY 350 AGAACTTCAGCCGCCAGGAAGTTGGCGCGGCCACCCTGCGGTTGTGGCGGTGTTGCCG 409
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 68 AGAACTTTAGTCGTCAAGAAGTCGGACTGCGGCTGTTGTTGGCGCTCTTGCTG 127

QY 410 GCCAGCTGTTCCGTTAAGGCTCTGGTGTGTGTCGATGGCCATCGCGGTTCGTTCTGG 469
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 128 GCCAGCTGTTTGGTTAAGGTAGTGGTCGAGTTGCAATGGCCATTGGTGGTCTCTTTTGG 187

QY 470 CGCGTCTGATTGGCTCTAAAATCGGTCAGAGCATGGACGACGAGATAAAATCAAATGA 529
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 188 GTGGAATTAATGGTTCTTAAATCGGTCAATCGATGGATCAGCAGATAAATAANGCTAA 247

QY 530 ACCAGTCTCTGGAAAAAGTGAAGCCGCCAGGTACTCGTTGGCGTAAATCCGGACACCG 589
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 248 ACCAGAGTTTGGAAAAAGTAAAAGCAGGGCAAGTGACAGCTTGGCGTAAATCCAGATACAG 307

QY 590 GTACAGCTACTCTGTGGACCCGGTTGCGACCTTACAGGCTTTACAACAACAGAGAACGCC 649
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 308 GCAATAGTTATAGTGTGTTGAGCCAGTCGGTACTTACCAGCGTTTACAATAAGCAAGCGTC 367

QY 650 GTCAGCAGTACTGCGCGCAATTTACAGCAGAAAGCCATGATCGCAGGTTCAGAAACAGGAAA 709
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 368 GCCAGCAATATTGCGAGAAATTCAGCAAAAGGCGATGATGGAGGCGCAAGCAAGAGA 427

QY 710 TCTACGCCACCGGTGCCCTCAGCCGATGCCGCTGGCAGGTGATTAGCACCCGAAAAA 768
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 428 TTTCAGCACTGCATGCCGCAACCGGATGTCGTTGGCAAGTCATTTCACACAGAAAAA 486
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 8
AAD11043
ID AAD11043 standard; DNA; 573 BP.
XX
XX AAD11043;
AC
XX AC
XX
XX 24-SEP-2001 (first entry)
XX
XX Clostridium cellulovorans cellulose binding domain-180 (CBD-180) DNA.
```

KW Polysaccharide modification: polysaccharide binding domain; PBD; paper;
 KW yarn; fiber; textile; biological crosslinker; mechanical property;
 KW Clostridium cellulovorans cellulose binding domain; wet strength;
 KW durability; elasticity; CBDelos; cellulose binding protein A; CBP A;
 KW CBD-180; ds.
 XX
 XX Clostridium cellulovorans.
 XX
 FH Key Location/Qualifiers
 FT CDS 3..572
 FT /*tag= a
 FT /product= "Cellulose binding domain-180 protein"
 FT /note= "CDS does not include stop codon"
 FT /partial
 XX
 XX W0200134091-A2.
 XX
 XX 17-MAY-2001.
 XX
 XX 02-NOV-2000; 2000WO-IL00708.
 XX
 XX 08-NOV-1999; 99US-0164140.
 XX 18-NOV-1999; 99US-0166389.
 XX
 XX (CBDT-) CBD TECHNOLOGIES LTD
 XX (YISS) YISSUM RES DEV CO HEBREW UNIV JERUSALEM.
 XX
 XX Levy I, Nussinovitch A, Shoseyov O;
 XX WPI: 2001-457121/49.
 XX P-PSDB; AAE05746.
 XX
 XX Preparation of a polysaccharide containing material having at least one
 XX desired structural, chemical, physical, electrical and/or mechanical
 XX property -
 XX
 XX Example 1.2; Fig 1e-1g; 121pp; English.
 XX
 XX The present invention relates to methods and compositions for cross-
 XX linking and/or modifying the properties of polysaccharide materials.
 XX The method involves treating the polysaccharide structure with a
 XX polysaccharide binding domain (PBD) fusion protein. The method is
 XX used to alter the structural, chemical, physical, electrical and
 XX mechanical properties of polysaccharide materials such as paper,
 XX yarns, fibers and textiles, using biological crosslinking agents.
 XX The polysaccharide containing materials have improved mechanical
 XX properties such as wet strengths, durability and elasticity. The PBD
 XX reagent is applied in the forming stage in fluting paper manufacture
 XX which eliminates the sizing step. The use of a biological crosslinker
 XX improves the recyclability of paper products. The PBD reagent maintains
 XX the fine fibers in a slurry therefore resulting in better recovery of
 XX raw materials. The PBD molecules are eluted by strong alkaline conditions
 XX which enhances the ability of the alkaline glue used in binding to
 XX penetrate paper.
 XX The present sequence is a DNA encoding Clostridium cellulovorans
 XX cellulose binding domain (CBD)-180. CBD-180 contains 180 amino acids of
 XX Clostridium cellulovorans CBD (CBDelos) of cellulose binding protein A
 XX (CBP A).
 XX
 XX Sequence 573 BP; 215 A; 112 C; 98 G; 148 T; 0 other;

Query Match 35.4%; Score 271.8; DB 22; Length 573;
 Best Local Similarity 99.3%; Pred. No. 2.8e-69;
 Matches 273; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ATGTCAGTTGAATCTCAACTTAACAAATCAGCACAAACAACTCAATACACCAATA 60
 DB 3 ATGTCAGTTGAATCTCAACTCTCAACAAATCAGCACAAACAACTCAATACACCAATA 62
 QY 61 ATCAAAATTTACTACACATCTGACAGTCAATTTAAATGACGTAAAGTTAGATAT 120
 DB 63 ATCAAAATTTACTACACATCTGACAGTCAATTTAAATGACGTAAAGTTAGATAT 122

QY 121 TATTACACAAGTGTGTACACAGGACAACACTTCTGGTGTGACCATGCTGGTGCAATTA 180
 DB 123 TATTACACAAGTGTGTACACAGGACAACACTTCTGGTGTGACCATGCTGGTGCAATTA 182
 QY 181 TTAGGAAATAGCTATGTTGATACACACTAGCAAAAGTGACAGCAAACTTCGTTAAAGAAACA 240
 DB 183 TTAGGAAATAGCTATGTTGATACACACTAGCAAAAGTGACAGCAAACTTCGTTAAAGAAACA 242
 QY 241 GCAAGCCCAACATCAACCTATGATACATATCTGTGA 275
 DB 243 GCAAGCCCAACATCAACCTATGATACATATCTGTGA 277
 RESULT 9
 AAD11044
 ID AAD11044 standard; DNA; 1030 BP.
 XX
 XX AAD11044;
 XX
 XX 24-SEP-2001 (first entry)
 XX
 XX Clostridium cellulovorans CBD cross linker protein (CCP) encoding DNA.
 XX
 XX Polysaccharide modification: polysaccharide binding domain; PBD; paper;
 XX yarn; fiber; textile; biological crosslinker; mechanical property;
 XX wet strength; durability; elasticity; cellulose binding domain; CBD;
 XX CBD cross linker protein; CCP; ds.
 XX
 XX Clostridium cellulovorans.
 XX
 XX Key Location/Qualifiers
 XX CDS 3..1028
 XX /*tag= a
 XX /product= "CCP protein"
 XX /transl_except= "(pos: 1020..1028, aa:Pro-Asp)"
 XX /note= "CDS does not include stop codon"
 XX /partial
 XX
 XX W0200134091-A2.
 XX
 XX 17-MAY-2001.
 XX
 XX 02-NOV-2000; 2000WO-IL00708.
 XX
 XX 08-NOV-1999; 99US-0164140.
 XX 18-NOV-1999; 99US-0166389.
 XX
 XX (CBDT-) CBD TECHNOLOGIES LTD
 XX (YISS) YISSUM RES DEV CO HEBREW UNIV JERUSALEM.
 XX
 XX Levy I, Nussinovitch A, Shoseyov O;
 XX WPI: 2001-457121/49.
 XX P-PSDB; AAE05747.
 XX
 XX Preparation of a polysaccharide containing material having at least one
 XX desired structural, chemical, physical, electrical and/or mechanical
 XX property -
 XX
 XX Example 1.2; Fig 2b-2e; 121pp; English.
 XX
 XX The present invention relates to methods and compositions for cross-
 XX linking and/or modifying the properties of polysaccharide materials.
 XX The method involves treating the polysaccharide structure with a
 XX polysaccharide binding domain (PBD) fusion protein. The method is
 XX used to alter the structural, chemical, physical, electrical and
 XX mechanical properties of polysaccharide materials such as paper,
 XX yarns, fibers and textiles, using biological crosslinking agents.
 XX The polysaccharide containing materials have improved mechanical
 XX properties such as wet strengths, durability and elasticity. The PBD
 XX reagent is applied in the forming stage in fluting paper manufacture
 XX which eliminates the sizing step. The use of a biological crosslinker
 XX improves the recyclability of paper products. The PBD reagent maintains

CC the fine fibers in a slurry therefore resulting in better recovery of
 CC raw materials. The PBD molecules are eluted by strong alkaline conditions
 CC which enhances the ability of the alkaline glue used in binding to
 CC penetrate paper.
 CC The present sequence is a DNA encoding Clostridium cellulovorans
 CC cellulose binding domain (CBD) cross linker protein (CCP).
 XX
 XX

Sequence 1030 BP; 402 A; 195 C; 166 G; 267 T; 0 other;

Query Match 35.4%; Score 271.8; DB 22; Length 1030;
 Best Local Similarity 99.3%; Pred. No. 3.8e-69;
 Matches 273; Conservative 0; Mismatches 2; Indels 0; Gaps* 0;

Oy 1 ATGTCAGTTGAATTTACAACTCTAACAAATCAGCACAAACAACTCAATTACACCAATA 60
 |||||
 Db 3 ATGTCAGTTGAATTTACAACTCTAACAAATCAGCACAAACAACTCAATTACACCAATA 62

Oy 61 ATCAAAATTAACACATCTGACAGTGAATTTAAATTTAAATGACGCTAAAGTTAGATAT 120
 |||||
 Db 63 ATCAAAATTAACACATCTGACAGTGAATTTAAATTTAAATGACGCTAAAGTTAGATAT 122

Oy 121 TATTACAAAGTGATGGTACAAAGCAAACTTTCTGGTGGACATGCTGGTGCAATTA 180
 |||||
 Db 123 TATTACAAAGTGATGGTACAAAGCAAACTTTCTGGTGGACATGCTGGTGCAATTA 182

Oy 181 TTAGGAAATAGCTATGTTGATAAATCAGCAAGTACAGCAAACTTCGTTAAAGAAACA 240
 |||||
 Db 183 TTAGGAAATAGCTATGTTGATAAATCAGCAAGTACAGCAAACTTCGTTAAAGAAACA 242

Oy 241 GCAAGCCCAACATCAACCTATGATACATATCTGGA 275
 |||||
 Db 243 GCAAGCCCAACATCAACCTATGATACATATGTTGA 277

RESULT 10

AAQ72917
 ID AAQ72917 standard; DNA; 486 BP.

XX
 AC AAQ72917;

XX 05-JUN-1995 (first entry)

XX Cellulose binding domain.

XX Cellulose binding domain; CBD; ds.

XX Clostridium cellulovorans.

XX Key Location/Qualifiers
 XX CDS 1..486
 XX /*tag= a

XX W09424158-A.

XX 27-OCT-1994.

XX 14-APR-1994; 94WO-US04132.

XX 14-APR-1993; 93US-0048164.

XX (RECC) UNIV CALIFORNIA.

XX (YISS) YISSUM RES & DEV CO.

XX Doi RH, Goldstein MA, Shoseyov O, Shpiegl I;

XX WPI; 1994-341767/42.

XX P-PSDB; AAR62634.

XX Isolated cellulose binding domain and fusion proteins - with
 PT applications. Including drug delivery, affinity separations. and
 PT diagnostic techniques

PS Claim 10; Fig 1; 125pp; English.

XX AAQ72917/R63634 is a novel isolated cellulose binding domain. It pref.
 CC binds crystalline with a Kd ranging from 1.5 microm - 0.5 microm,
 CC pref. with a Kd of <1.2 microm, esp. <1.0 microm. The protein and
 CC the encoding DNA are claimed, as is a CBD fusion protein (FP)
 CC comprising the CBD and a second protein. The second protein is pref.
 CC Protein A, heparinase, a hormone or an enzyme capable of degrading
 CC an environmental pollutant, or an HSP, HSP antibody, cross-reactive
 CC HSP-related protein or an antigenic portion of this. The CBD and FP
 CC may be used in drug delivery, affinity seps. and diagnostic
 CC techniques. CBD nucleic acid may be obtd. from a variety of cell
 CC sources that produce CBDs that bind with high affinity and in a
 CC reversible manner or that produce CBD encoding mRNA. The preferred
 CC source of CBD encoding nucleic acid is C. cellulovorans.
 XX

Sequence 486 BP; 194 A; 89 C; 79 G; 124 T; 0 other;

Query Match 35.2%; Score 270.2; DB 15; Length 486;
 Best Local Similarity 98.9%; Pred. No. 7.4e-69;
 Matches 272; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Oy 1 ATGTCAGTTGAATTTACAACTCTAACAAATCAGCACAAACAACTCAATTACACCAATA 60
 |||||

Db 16 ATGTCAGTTGAATTTACAACTCTAACAAATCAGCACAAACAACTCAATTACACCAATA 75

Oy 61 ATCAAAATTAACACATCTGACAGTGAATTTAAATTTAAATGACGCTAAAGTTAGATAT 120
 |||||

Db 76 ATCAAAATTAACACATCTGACAGTGAATTTAAATTTAAATGACGCTAAAGTTAGATAT 135

Oy 121 TATTACAAAGTGATGGTACAAAGCAAACTTTCTGGTGGACATGCTGGTGCAATTA 180
 |||||

Db 136 TATTACAAAGTGATGGTACAAAGCAAACTTTCTGGTGGACATGCTGGTGCAATTA 195

Oy 181 TTAGGAAATAGCTATGTTGATAAATCAGCAAGTACAGCAAACTTCGTTAAAGAAACA 240
 |||||

Db 196 TTAGGAAATAGCTATGTTGATAAATCAGCAAGTACAGCAAACTTCGTTAAAGAAACA 255

Oy 241 GCAAGCCCAACATCAACCTATGATACATATCTGGA 275
 |||||

Db 256 GCAAGCCCAACATCAACCTATGATACATATGTTGA 290

RESULT 11

AAV74072
 ID AAV74072 standard; DNA; 486 BP.

XX
 AC AAV74072;

XX 09-MAR-1999 (first entry)

XX C. cellulovorans CbpA CBD DNA.

XX Cellulose binding domain; CBD; CbpA; fusion protein; diagnosis; HSP;
 KW immunosassay; heat-shock protein; cross reactive protein; detection;
 KW antigenic fragment; antibody; insulin-dependent diabetes mellitus;
 KW cellulose; chitin; cellulolytic; amorphogenic; ds.

XX Clostridium cellulovorans.

XX Key Location/Qualifiers
 XX CDS 1..486
 XX /*tag= a
 XX /*product= "cellulose binding domain"

XX US5856201-A.

XX 05-JAN-1999.

XX 27-OCT-1994; 94US-0330394.

XX 27-OCT-1994; 94US-0330394.

XX 14-APR-1993; 93US-0048164.

XX 14-APR-1994; 94WO-US04132.

XX (REGC) UNIV CALIFORNIA.
 PA (YISS) YISSUM RES & DEV CO.
 XX
 PI Doi RH, Goldstein MA, Shoseyov O, Shpiegl I;
 XX
 XX WPI: 1999-105130/09.
 DR P-PSDB; AAW90077.
 XX
 XX Detection of a specific analyte by reaction with binding agent fused
 PT to cellulose binding domain - and subsequent treatment with
 PT cellulose and reaction of insoluble product with a label specific
 PT for the analyte
 XX
 XX Example 7.2.1; Fig 1A-B; 63pp; English.
 PS
 XX This sequence encodes a cellulose binding domain (CBD) derived from the
 CC CbpA protein of Clostridium cellulovorans. The sequence is used the
 CC construction of a fusion protein which can be used in diagnostic
 CC immunoassays, e.g. to detect heat-shock proteins (HSP) and their
 CC cross-reactive proteins, antigenic fragments or HSP-specific antibodies
 CC (which indicate insulin-dependent diabetes mellitus, or susceptibility
 CC to it. The CBD, has very high affinity for cellulose (including
 CC crystalline forms) and chitin (dissociation constant 0.8-1.4 mu M), but
 CC has almost no cellulolytic or amorphogenic activities. The CBD binds over
 CC a wide pH range and is not released from cellulose by washing with water.
 XX
 SQ Sequence 486 BP; 194 A; 89 C; 79 G; 124 T; 0 other;
 Query Match 35.2%; Score 270.2; DB 20; Length 486;
 Best Local Similarity 98.9%; Pred. No. 7.4e-69;
 Matches 272; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 QY 1 ATGTCAGTTGAATTCACAACTCTAACAAATCAGCACAACAACTCAATTACACCAATA 60
 DB 16 ATGTCAGTTGAATTTACAACTCTAACAAATCAGCACAACAACTCAATTACACCAATA 75
 QY 61 ATCAAAATTTACTAACACATCTGCAGTGATTTAAATTTAAATGACGTAAGTTAGATAT 120
 DB 76 ATCAAAATTTACTAACACATCTGCAGTGATTTAAATTTAAATGACGTAAGTTAGATAT 135
 QY 121 TATTACACAAGTGTGATACACAGGACAACTTCTGTTGTCACCATGCTGGTGCAATTA 180
 DB 136 TATTACACAAGTGTGATACACAGGACAACTTCTGTTGTCACCATGCTGGTGCAATTA 195
 QY 181 TTAGGAATAGCTATGTTGATACACTAGCAAAAGTGACAGCAAACTTCGTTAAAGAAACA 240
 DB 196 TTAGGAATAGCTATGTTGATACACTAGCAAAAGTGACAGCAAACTTCGTTAAAGAAACA 255
 QY 241 GCAAGCCCAACATCAACCTATGATACATATCTGGA 275
 DB 256 GCAAGCCCAACATCAACCTATGATACATATGTTGA 290
 RESULT 12
 AAX24930
 ID AAX24930 standard; DNA: 499 BP.
 XX
 XX AAX24930;
 XX
 XX 21-JUN-1999 (first entry)
 DT
 XX Clostridium cellulovorans cbpA cellulose binding domain DNA.
 DE
 XX Cellulose binding domain; CBD: cbpA; endo-1,4-beta-glucanase; Cell;
 KW Arabidopsis thaliana; transgenic plant; crop improvement;
 KW morphology; cell wall; ds.
 XX
 OS Clostridium cellulovorans.
 XX
 PN WO9907830-A1.
 XX
 PD 18-FEB-1999.

XX 26-JUL-1998; 98WO-IL00345.
 PF
 XX 13-JAN-1998; 98US-0006636.
 PR
 PR 27-JUL-1997; 97IL-0121404.
 PR
 PR 13-JAN-1998; 98US-0006632.
 XX
 PA (YISS) YISSUM RES & DEV CO.
 XX
 XX Shani Z, Shoseyov O, Shpiegl E;
 PI
 XX WPI: 1999-180488/15.
 DR
 XX Transgenic plants expressing cell-wall modulating protein - have
 PT altered morphology, e.g. increased growth, modified fiber length or
 PT cellulose content
 PT
 XX Disclosure; Page 134; 144pp; English.
 PS
 XX This DNA fragment encodes the cellulose binding domain (CBD) of
 CC the cbpA protein of Clostridium cellulovorans. It was obtained
 CC by PCR amplification (see also AAX24952-53), and was used in the
 CC construction of binary vector pCC1, in which cbd was joined to
 CC the promoter and signal region of the novel endo-1,4-beta-glucanase
 CC cell gene (see AAX24923) of Arabidopsis thaliana. Expression of cbd
 CC in transgenic tobacco plants modulated their growth. The cell
 CC promoter can provide expression of any protein in elongating
 CC tissue. CBD is an example of a cell wall modulation transgene
 CC used to alter the structure or morphology of a plant. Transgenic
 CC plants of the invention may have altered biomass, growth, yield,
 CC greater or less resistance to biodegradation, be more or less
 CC digestible by ruminants, have altered cellulose content, larger or
 CC smaller leaves, etc., when compared to non-transgenic plants of the
 CC same species.
 XX
 SQ Sequence 499 BP; 197 A; 93 C; 82 G; 127 T; 0 other;
 Query Match 35.2%; Score 270.2; DB 20; Length 499;
 Best Local Similarity 98.9%; Pred. No. 7.5e-69;
 Matches 272; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 QY 1 ATGTCAGTTGAATTCACAACTCTAACAAATCAGCACAACAACTCAATTACACCAATA 60
 DB 21 ATGTCAGTTGAATTTACAACTCTAACAAATCAGCACAACAACTCAATTACACCAATA 80
 QY 61 ATCAAAATTTACTAACACATCTGCAGTGATTTAAATTTAAATGACGTAAGTTAGATAT 120
 DB 81 ATCAAAATTTACTAACACATCTGCAGTGATTTAAATTTAAATGACGTAAGTTAGATAT 140
 QY 121 TATTACACAAGTGTGATACACAAGGACAACTTCTGTTGTCACCATGCTGGTGCAATTA 180
 DB 141 TATTACACAAGTGTGATACACAAGGACAACTTCTGTTGTCACCATGCTGGTGCAATTA 200
 QY 181 TTAGGAATAGCTATGTTGATACACTAGCAAAAGTGACAGCAAACTTCGTTAAAGAAACA 240
 DB 201 TTAGGAATAGCTATGTTGATACACTAGCAAAAGTGACAGCAAACTTCGTTAAAGAAACA 260
 QY 241 GCAAGCCCAACATCAACCTATGATACATATCTGGA 275
 DB 261 GCAAGCCCAACATCAACCTATGATACATATGTTGA 295
 RESULT 13
 AAD11042
 ID AAD11042 standard; DNA: 507 BP.
 XX
 AC AAD11042;
 XX
 XX 24-SEP-2001 (first entry)
 DT
 XX Clostridium cellulovorans cellulose binding domain (CBD:cbpA) DNA.
 DE Polysaccharide modification; polysaccharide binding domain; PBD; paper;
 XX

KW yarn; fiber; textile; biological crosslinker; mechanical property;
KW Clostridium cellulovorans cellulose binding domain; wet strength;
KW durability; elasticity; CBDclo; cellulose binding protein A; CBP A; ds.
XX
XX Clostridium cellulovorans.
XX
XX Key Location/Qualifiers
FT CDS 3..494
FT /tag= a
FT /product= "CBDclo protein"
XX
XX WO200134091-A2.
XX
XX 17-MAY-2001.
XX
XX 02-NOV-2000; 2000WO-IL00708.
XX
XX 08-NOV-1999; 99US-0164140.
PR 18-NOV-1999; 99US-0166389.
XX
XX (CBDT-) CBD TECHNOLOGIES LTD.
PA (YISS) YISSUM RES DEV CO HEBREW UNIV JERUSALEM.
XX
XX Levy I, Nussinovitch A, Shoseyov O;
XX
XX WPI; 2001-457121/49.
DR P-PSDB; AAE05745.
XX
XX Preparation of a polysaccharide containing material having at least one
PT desired structural, chemical, physical, electrical and/or mechanical
PT property -
XX
XX Example 1.1; Page 111; 121pp; English.
XX
XX The present invention relates to methods and compositions for cross-
CC linking and/or modifying the properties of polysaccharide materials.
CC The method involves treating the polysaccharide structure with a
CC polysaccharide binding domain (PBD) fusion protein. The method is
CC used to alter the structural, chemical, physical, electrical and
CC mechanical properties of polysaccharide materials such as paper,
CC yarns, fibers and textiles, using biological crosslinking agents.
CC The polysaccharide containing materials have improved mechanical
CC properties such as wet strengths, durability and elasticity. The PBD
CC reagent is applied in the forming stage in fluting paper manufacture
CC which eliminates the sizing step. The use of a biological crosslinker
CC improves the recyclability of paper products. The PBD reagent maintains
CC the fine fibers in a slurry therefore resulting in better recovery of
CC raw materials. The PBD molecules are eluted by strong alkaline conditions
CC which enhances the ability of the alkaline glue used in binding to
CC penetrate paper.
CC The present sequence is a DNA encoding Clostridium cellulovorans
CC cellulose binding domain (CBDclo) of cellulose binding protein A
CC (CBP A).
XX
XX Sequence 507 BP; 200 A; 94 C; 84 G; 129 T; 0 other;
Query Match 35.2%; Score 270.2; DB 22; Length 507;
Best Local Similarity 98.9%; Pred. No. 7.6e-69;
Matches 272; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 ATGTCAGTTGAATCTTACAACTCTAACAAATCAGCAGCAACAACTCAATTACACCAATA 60
DB 21 ATGTCAGTTGAATTTTAACTCTAACAAATCAGCAGCAACAACTCAATTACACCAATA 80
QY 61 ATCAAATTAATACACATCTGACAGTGATTTAAATTTAAATGAGCTAAAAGTTAGATAT 120
DB 81 ATCAAATTAATACACATCTGACAGTGATTTAAATTTAAATGAGCTAAAAGTTAGATAT 140
QY 121 TATTACAAAGTGATGGTACACAGGCAAACTTTCTGGTGTGACCATGCTGGTGCAATA 180
DB 141 TATTACAAAGTGATGGTACACAGGCAAACTTTCTGGTGTGACCATGCTGGTGCAATA 200
QY 161 TTAGGAATAGTATGTTGTAATACACTAGCAAGACGACAGCAAACTTCGTTAAAGAAACA 240

DB 201 TTAGGAATAGTATGTTGTAATACACTAGCAAGACGACAGCAAACTTCGTTAAAGAAACA 260
QY 241 GCAAGCCCAACATCAACCTATGATACATATCTGGA 275
DB 261 GCAAGCCCAACATCAACCTATGATACATATCTGGA 295
RESULT 14
AAD11046
ID AAD11046 standard; DNA; 984 BP.
XX
XX AAD11046;
XX
XX 24-SEP-2001 (first entry)
XX
XX Chimeric S peptide-cellulose binding domain-S protein encoding DNA.
XX
XX Polysaccharide modification; polysaccharide binding domain; PBD; paper;
KW yarn; fiber; textile; biological crosslinker; mechanical property;
KW wet strength; durability; elasticity; cellulose binding domain; CBD;
KW Chimeric protein; S peptide-cellulose binding domain-S protein;
KW Speg-CBD-Sprot; bovine; ds.
XX
XX Chimeric - Clostridium cellulovorans.
OS Chimeric - Bos sp.
XX
XX Key Location/Qualifiers
FT CDS 1..984
FT /tag= a
FT /product= "Speg-CBD-Sprot protein"
FT /trans_except= "(pos: 979..981, aa:Xaa)"
FT /note= "Xaa corresponds to in-frame stop codon;
FT CDS does not include start and stop codon"
FT /partial
FT misc_feature 68..624
FT /tag= b
FT /note= "This region is derived from C. cellulovorans"
FT misc_feature 652..981
FT /tag= c
FT /note= "This region is derived from bovine"
XX
XX WO200134091-A2.
XX
XX 17-MAY-2001.
XX
XX 02-NOV-2000; 2000WO-IL00708.
XX
XX 08-NOV-1999; 99US-0164140.
PR 18-NOV-1999; 99US-0166389.
XX
XX (CBDT-) CBD TECHNOLOGIES LTD.
PA (YISS) YISSUM RES DEV CO HEBREW UNIV JERUSALEM.
XX
XX Levy I, Nussinovitch A, Shoseyov O;
XX
XX WPI; 2001-457121/49.
DR P-PSDB; AAE05749.
XX
XX Preparation of a polysaccharide containing material having at least one
PT desired structural, chemical, physical, electrical and/or mechanical
PT property -
XX
XX Example 4; Fig 4b-4g; 121pp; English.
XX
XX The present invention relates to methods and compositions for cross-
CC linking and/or modifying the properties of polysaccharide materials.
CC The method involves treating the polysaccharide structure with a
CC polysaccharide binding domain (PBD) fusion protein. The method is
CC used to alter the structural, chemical, physical, electrical and
CC mechanical properties of polysaccharide materials such as paper,
CC yarns, fibers and textiles, using biological crosslinking agents.
CC The polysaccharide containing materials have improved mechanical

[illegible]

Search completed: February 22, 2003, 14:45:04
Job time : 221.439 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 22, 2003, 14:26:45 ; Search time 1523.18 Seconds
(without alignments)
8165.910 Million cell updates/sec

Title: US-09-677-374-5
Perfect score: 768
Sequence: 1 atgtcagttgaattctacaa.....agggtattagaccgaaaaaa 768

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

- EST:*
- 1: em_estba:*
 - 2: em_esthum:*
 - 3: em_estin:*
 - 4: em_estmu:*
 - 5: em_estov:*
 - 6: em_estpl:*
 - 7: em_estro:*
 - 8: em_htc:*
 - 9: gb_est1:*
 - 10: gb_est2:*
 - 11: gb_htc:*
 - 12: gb_est3:*
 - 13: gb_est4:*
 - 14: gb_est5:*
 - 15: em_estfun:*
 - 16: em_estom:*
 - 17: gb_gss:*
 - 18: em_gss_hum:*
 - 19: em_gss_inv:*
 - 20: em_gss_pln:*
 - 21: em_gss_vrt:*
 - 22: em_gss_fun:*
 - 23: em_gss_mem:*
 - 24: em_gss_mus:*
 - 25: em_gss_other:*
 - 26: em_gss_pro:*
 - 27: em_gss_rod:*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|-------------|
| c 1 | 47 | 6.1 | 582 | 13 | BJ429910 |
| 2 | 43.4 | 5.7 | 551 | 9 | AU037653 |
| 3 | 43 | 5.6 | 1100 | 17 | CNS00FCD |
| 4 | 42.4 | 5.5 | 470 | 17 | AZ799648 |
| 5 | 42.2 | 5.5 | 976 | 17 | CNS04E5M |
| 6 | 42 | 5.5 | 238 | 12 | BG240789 |

| | | | | | |
|----|-----|-----|------|----------|------------|
| 7 | 5.5 | 42 | 17 | AQ260501 | CITBI-EI- |
| 8 | 5.5 | 402 | 12 | BG356895 | BG356895 |
| 9 | 5.5 | 419 | 10 | BE355894 | DG1_11_E1 |
| 10 | 5.5 | 428 | 12 | BF176742 | BF176742 |
| 11 | 5.5 | 446 | 12 | BG052290 | EM1_4_F01 |
| 12 | 5.5 | 500 | 12 | BF586874 | BG052290 |
| 13 | 5.5 | 507 | 12 | BG053591 | BF586874 |
| 14 | 5.5 | 531 | 12 | BG053014 | BG053591 |
| 15 | 5.5 | 537 | 12 | BG713844 | RHI22_11- |
| 16 | 5.5 | 561 | 12 | BF587805 | BG053014 |
| 17 | 5.5 | 577 | 12 | BG673858 | EM1_20_C0 |
| 18 | 5.5 | 582 | 10 | BE356763 | BG673858 |
| 19 | 5.5 | 591 | 10 | BE356800 | EM1_20_C0 |
| 20 | 5.5 | 597 | 12 | BG102589 | DG1_12_E1 |
| 21 | 5.5 | 598 | 10 | AW672446 | BE356800 |
| 22 | 5.5 | 600 | 10 | BE360868 | BG102589 |
| 23 | 5.5 | 620 | 12 | BF481524 | LG1_359_D |
| 24 | 5.5 | 649 | 10 | BE355895 | DG1_67_A0 |
| 25 | 5.5 | 649 | 9 | AA352399 | BF481524 |
| 26 | 5.5 | 649 | 13 | BM031893 | DG1_11_E1 |
| 27 | 5.5 | 649 | 10 | AW021656 | AA352399 |
| 28 | 5.5 | 649 | 9 | AA663164 | EST60642 |
| 29 | 5.5 | 649 | 13 | BI493124 | BM031893 |
| 30 | 5.5 | 649 | 17 | AZ447316 | AA663164 |
| 31 | 5.5 | 649 | 10 | BE395015 | BI493124 |
| 32 | 5.5 | 649 | 13 | BI492596 | AZ447316 |
| 33 | 5.5 | 649 | 10 | AW328753 | BE395015 |
| 34 | 5.5 | 649 | 742 | BF792098 | BI492596 |
| 35 | 5.5 | 649 | 878 | BM449796 | AW328753 |
| 36 | 5.5 | 649 | 958 | CNS001FF | BF792098 |
| 37 | 5.5 | 649 | 1058 | BM478067 | BM449796 |
| 38 | 5.5 | 649 | 1079 | BF792265 | AL074914 |
| 39 | 5.5 | 649 | 1196 | BM542460 | Drosophill |
| 40 | 5.5 | 649 | 1201 | CNS0167M | BM478067 |
| 41 | 5.5 | 649 | 702 | W27594 | BF792265 |
| 42 | 5.5 | 649 | 737 | BM166352 | BM542460 |
| 43 | 5.5 | 649 | 928 | CNS00DKY | AL106396 |
| 44 | 5.5 | 649 | 1249 | BM909360 | Drosophill |
| 45 | 5.5 | 649 | 942 | CNS018GS | W27594 |

ALIGNMENTS

RESULT 1
BJ429910/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

BJ429910 Dictyostellium discoidium linear EST 13-MAR-2002
Dictyostellium discoidium cDNA library, VF Dictyostellium
Dictyostellium discoidium cDNA clone ddv5109 3', mRNA sequence.
BJ429910 GI:19404632
Dictyostellium discoidium.
Dictyostellium discoidium.
Eukaryota; Mycetozoa; Dictyostelid; Dictyostellid; Dictyostellium.
1 (bases 1 to 582)
Urushihara, H., Tanaka, Y., Kohara, Y. and Shin-i, T.
Full length cDNA of Dictyostellium discoidium at the vegetative stage
Unpublished (2002)
Contact: Tadasu Shin-i
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshini@genes.nig.ac.jp.
Location/Qualifiers
1. 582
/organism="Dictyostellium discoidium"
/strain="AX4"
/db_xref="taxon:44689"
/clone="ddv5109"

FEATURES
source

```
/clone_lib="Dictyostelium discoideum cdna library, VF"
/sex="mat A"
/dev_stage="Growth phase"
BASE COUNT 116 a 59 c 320 t 3 others
ORIGIN

Query Match 6.1%; Score 47; DB 13; Length 582;
Best Local Similarity 47.7%; Pred. No. 0.052;
Matches 137; Conservative 0; Mismatches 150; Indels 0; Gaps 0;

QY 10 GAATTTCTACAACCTTAACAAATACGACACAAACAACTCAATACACCAATATCAAAAT 69
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 435 GAATCAATACACTACAGATCACTACACTACTACTACTACTACTACTACTACTACTACT 376
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 70 ACTACACATCTGACAGTGATTTAAATTAATGACGTAAGTTAGATATATTATACACA 129
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 375 ACAACAACACTGCAGATAACCAAGAAAGAAAGAAATATAAAATATAAAATCA 316
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 130 AGTGATGTACACAGGACAACTTCTGGTGTGACCATCTGGTGCATTATTAGGAAT 189
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 315 AATACACTACAGATCACTACACTACTACTACTACTACTACTACTACTACTACTACT 256
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 190 AGCTATGTGATAACACTAGCAAGTACGACGAACTTCGTTTAAAGAAACAGCAAGCCCA 249
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 255 ACAACTGCAGATAACCAAGAAAGAAAGAAATATAAAATATAAAATCAAAACCA 196
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 250 ACATCAACTATGATACATATCTGGATCGTCTCATATCGGTGGTGG 296
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 195 ATATCAACTGATAATAAAATATAGACTCAACAACACTACTACTAGTGTG 149
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 2
LOCUS AU037653 551 bp mRNA linear EST 29-MAR-1999
DEFINITION AU037653 Dictyostelium discoideum SS (H.Urushihara) Dictyostellium
ACCESSION AU037653
VERSION AU037653.1 GI:3984406
KEYWORDS EST.
SOURCE Dictyostelium discoideum.
ORGANISM Dictyostelium discoideum
REFERENCE Dictyostellium discoideum.
AUTHORS Eukaryota; Mycetozoa; Dictyostellida; Dictyostellium.
Moriyama, T., Urushihara, H., Saito, T., Ugawa, Y., Mizuno, H., Yoshida, M.,
Yoshino, R., Mitra, B. N., Pi, M., Sato, T., Takemoto, K., Yasukawa, H.,
Williams, J., Maeda, M., Takeuchi, I., Ochiai, H. and Tanaka, Y.
The Dictyostelium developmental cDNA project: generation and
analysis of expressed sequence tags from the first-finger stage of
development
DNA Res. 5 (6), 335-340 (1998)
JOURNAL 99156227
MEDLINE Contact: Hideko Urushihara
COMMENT Institute of Biological Sciences
University of Tsukuba
1-1-1 Tennoudai, Tsukuba, Ibaraki 305-8572, Japan
Tel: 81-298-53-4664
Fax: 81-298-53-6614
Email: hideko@biol.tsukuba.ac.jp
PROJECT = 'Dictyostelium discoideum cdna project in Japan'.
FEATURES
source 1. 551
Location/Qualifiers
/organism="Dictyostelium discoideum"
/strain="AX4"
/db_xref="taxon:44689"
/clone="SS138"
/clone_lib="Dictyostelium discoideum SS (H.Urushihara)"
/dev_stage="slug"
BASE COUNT 266 a 63 c 39 g 183 t
ORIGIN

Query Match 5.7%; Score 43.4; DB 9; Length 551;
Best Local Similarity 52.5%; Pred. No. 0.05;
Matches 95; Conservative 0; Mismatches 86; Indels 0; Gaps 0;
```

```
QY 32 CAGCACAACAACTCAATTAACCAATAATCAAAATTAATAACACATCTGACATGATT 91
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 34 CAACATATAACAAATAATATTAATCAATAAAATAATCAGTAATCAAAAAATTAAT 93
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 92 TAAATTTAAATGACGTAAAGTTAGATATTTACACAGTGTGATGATGATGATGATGAT 151
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 94 CAAATACATATTTTAAATAAAGAAATTTTAAATCAAAATGTTACTTCATCAAAATCAA 153
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 152 CTTTCTGGTGTGACCATCTGGTGTGATTTAGGAAATAGTATGTTGATACACATAGCA 211
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 154 ATTCAATTAATGTAATTGCAACTTTAATTTTACCACAAATTTGATGATATAATAATA 213
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 212 A 212
DB 214 A 214

RESULT 3
LOCUS CNS00FCD 1100 bp DNA linear GSS 03-JUN-1999
DEFINITION Drosophila melanogaster genome survey sequence TET3 end of BAC:
BACR31D18 of RPCI-98 library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
ACCESSION AL070512
VERSION AL070512.1 GI:4950453
KEYWORDS GSS.
SOURCE Drosophila melanogaster.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
REFERENCE Genoscope.
AUTHORS Direct Submission
TITLE Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
JOURNAL BP 191 91006 EVRY cedex - FRANCE (E-mail : segref@genoscope.cns.fr
COMMENT - Web : www.genoscope.cns.fr)
Determination of this BAC-end sequence was carried out as part of a
collaboration with the Berkeley Drosophila Genome Project (BDGP).
The BDGP is constructing a physical map of the Drosophila
melanogaster genome using these BACs. For further information
please see http://www.fruitfly.org The BDGP Drosophila
melanogaster BAC library was prepared by Kazutoyo Osoegawa and
Aaron Mamoser in Pieter de Jong's laboratory in the Department of
Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
NY. The library is named RPCI-98 and was constructed by partial
ECORI digestion of Drosophila DNA provided by the BDGP from the
isogenic strain y2; cn bw sp, the same strain used for the BDGP's
P1 and EST libraries. A more detailed description of the library
and how to order individual BAC clones, the entire library, or
filters for hybridization from the BACPAC Resource Center can be
found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES
source 1. 1100
Location/Qualifiers
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone="BACR31D18"
/clone_lib="RPCI-98"
/note="end : TET3"
BASE COUNT 356 a 100 c 111 g 137 t 396 others
ORIGIN

Query Match 5.6%; Score 43; DB 17; Length 1100;
Best Local Similarity 33.7%; Pred. No. 0.82;
Matches 83; Conservative 6; Mismatches 157; Indels 0; Gaps 0;

QY 25 ACAAATCAGCAACAACTCAATTAACCAATAATCAAAATTAATAACACATCTGAC 84
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 104 AAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAA 163
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 85 AGTGATTTAAATTTAAATGACGTAAAGTTAGATATTTTACACAAAGTGTGATGATGAT 144
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

Db 164 CGASMGAAAAAANAAGGGAANNANNNAAAAAHHAAAYAAANATAAAWAAAAA 223
QY 145 GGCAAACTTCTGGTGTGACCATGCTGTGCATTATTAGGAATAGCTATGTTGATAAC 204
Db 224 CAAGACATTANNNNANTNNNTNNNNANNNNNANNNANNNNNANNAANN 283
QY 205 ACTAGCAAGTGACACCAAACTTCGTTAAAGAACAGCAGCCCAACATCAACCTATGAT 264
Db 284 NAAAAAANNNANNNANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 343
QY 265 ACATAT 270
Db 344 AAKAT 349

RESULT 4
AZ799648
LOCUS 470 bp DNA linear GSS 16-FEB-2001
DEFINITION 2M0057B14F Mouse 10kb plasmid UUGCLM library Mus musculus genomic
clone UUGC2M0057B14 F, DNA sequence.
ACCESSION AZ799648
VERSION AZ799648
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 470)
AUTHORS Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamil, C.,
Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly,
M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausern, A.,
and Wright, D., Weiss, R.
TITLE Mouse whole genome scaffolding with paired end reads from 10kb
plasmid inserts
JOURNAL Unpublished (2000)
COMMENT Contact: Robert B. Weiss
University of Utah Genome Center
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert length: 10000 Std Error: 0.00
Plate: 0057 row: B column: 14
Seq primer: CGTTGTAACACGACGCCAGT
Class: plasmid ends
High quality sequence stop: 470.
FEATURES
source Location/Qualifiers
1..470
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGC2M0057B14"
/clone_lib="Mouse 10kb plasmid UUGCLM library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/notes="Vector: PWD42nv; Purified genomic DNA from M.
musculus C57BL/6J (male) was obtained from the Jackson
Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA
was hydrodynamically sheared by repeated passage through a
0.005 inch orifice at constant velocity. The sheared DNA
was blunt end-repaired with T4 DNA polymerase and T4
polynucleotide kinase. Adaptor oligonucleotides were
ligated to the blunt ends in high molar excess. The
adapted DNA was purified and size-selected for a 9.5 to
10.5 kb range using preparative agarose gel
electrophoresis. Vector DNA was prepared from a derivative
of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
inducible derivative of plasmid R1. The vector was ligated
with adaptors complementary to the insert adaptors and
purified. The sheared, adapted mouse DNA was annealed to

adapted vector DNA, and transformed into
chemically-competent E. coli XL10-Gold (Stratagene) cells
and selected for ampicillin resistance."

BASE COUNT 138 a 97 c 132 g 103 t
ORIGIN

Query Match 5.5%; Score 42.4; DB 17; Length 470;
Best Local Similarity 57.6%; Pred. No. 0.89;
Matches 76; Conservative 0; Mismatches 56; Indels 0; Gaps 0;

QY 2 TGTCAAGTTGAATCTACAACTCTAACAAATCAGCACAAACAACTCAATACACCAATAA 61
Db 297 TGGTAGCTGTTGGTGTACATCTAGTTTACCACCAACCACTAAACCAACCAAAA 356
QY 62 TCAAAATTAACACACATCTGACAGTGATTTAAATTTAAATGAGTAAAAAGTTAGATATT 121
Db 357 TCAAACTCCAAATAACCCAGCATAAATGATGACCTCAATGTTGTTATTG 416
QY 122 ATTACACAATG 133
Db 417 ATGTCATTGGTG 428

RESULT 5
CNS04E5M/c
LOCUS 976 bp DNA linear GSS 21-MAY-2000
DEFINITION Tetraodon nigroviridis genome survey sequence T7 end of clone
103P02 of library G from Tetraodon nigroviridis, genomic survey
sequence.
AL286627
VERSION AL286627.1 GI:8025084
KEYWORDS GSS; genome survey sequence.
SOURCE Tetraodon nigroviridis.
ORGANISM Tetraodon nigroviridis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
Tetraodontidae; Tetraodon.
REFERENCE 1 (bases 1 to 976)
AUTHORS Roest-Crollius, H., Jaillon, O., Dasilva, C., Bouneau, L., Fisher, C.,
Bernot, A., Fizames, C., Wincker, P., Brottier, P., Quetier, F.,
Saurin, W. and Weissenbach, J.
TITLE Human gene number estimate provided by genome wide analysis using
Tetraodon nigroviridis DNA sequence
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 976)
AUTHORS Roest-Crollius, H., Jaillon, O., Dasilva, C., Fizames, C., Fisher, C.,
Bouneau, L., Billault, A., Quetier, F., Saurin, W., Bernot, A. and
Weissenbach, J.
TITLE Characterization and repeat analysis of the compact genome of the
freshwater pufferfish Tetraodon nigroviridis
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 976)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (12-APR-2000)
COMMENT This sequence is a single read and was generated as part of a large
scale clone-end sequencing project of the Tetraodon nigroviridis
genome. For more information, please take a look at
http://www.genoscope.cns.fr/Tetraodon.
FEATURES
source Location/Qualifiers
1..976
/organism="Tetraodon nigroviridis"
/db_xref="taxon:99883"
/clone="103P02"
/clone_lib="G"
/notes="Genoscope sequence ID : CORGI03DH01LP1-end : T7"

BASE COUNT 243 a 151 c 137 g 327 t 118 others
ORIGIN

Query Match 5.5%; Score 42.2; DB 17; Length 976;
Best Local Similarity 33.5%; Pred. No. 1.3;
Matches 116; Conservative 61; Mismatches 168; Indels 1; Gaps 1;

Email: mmpratt@uga.edu
Sequences have been trimmed to exclude PolyA, vector and regions below Phred quality 16. The threshold for highest quality sequence is 20.

Seq primer: PolyTMix
High quality sequence start: 32
High quality sequence stop: 427
POLYA-No.

FEATURES

Location/Qualifiers

1..428

/organism="Sorghum bicolor"

/db.xref="taxon:4558"

/clone.lib="Embryo 1 (EM1)"

/note="Organ: Embryos germinated for 24 hr; Vector:

pBluescript II from Lambda zap II; Site_1: XhoI; Site_2:

EcoRI; The library was made from poly-A RNA in the cloning

vector lambda ZAP II. Clones to be sequenced were

prepared by mass excision."

BASE COUNT 103 a 80 c 119 g 126 t

ORIGIN

Query Match 5.5%; Score 42; DB 12; Length 428;

Best Local Similarity 59.0%; Pred. No. 1.1;

Matches 72; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

QY 345 CGCCAGAACTTCAGCGCCAGAGTTGGCGCGCCAGCGGTGTGTGGCGGTGT 404

DB 58 CGCGAGAGAGCTGACCGAGGAGGTGCGACGAGATGATCCGTGAGGCTGACGTCGATGG 117

QY 405 TCCTCCGCCAGCTTTCGGTAAAGGCTCTGCTGTGTCGATGCCATCGCGCGGTGCGGT 464

DB 118 CGACGCCAGATCAACTATGAAGAGTTTGTAAAGTTATGATGGCCAAAGTGAGGAGCGGT 177

QY 465 TC 466

DB 178 CC 179

RESULT 11

LOCUS

BG052290 RH122_12_G04_g1_A003 Rhizome2 (RH122) Sorghum propinquum cDNA, mRNA

446 bp mRNA linear EST 25-JAN-2001

DEFINITION

sequence.

ACCESSION

BG052290

VERSION

BG052290.1 GI:12506830

KEYWORDS

EST.

SOURCE

Sorghum propinquum.

ORGANISM

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC

clade; Panicoideae; Andropogoneae; Sorghum.

REFERENCE

1 (bases 1 to 446)

AUTHORS

Cordonnier-Pratt,M.-M., Gingle,A., Sudman,A., Marsala,C. and Pratt

,L.H.

TITLE

An EST database from Sorghum: Sorghum propinquum rhizomes

JOURNAL

Unpublished (2000)

COMMENT

Contact: Cordonnier-Pratt MM

Laboratory for Genomics and Bioinformatics

The University of Georgia, Department of Plant Biology

Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA

Tel: 706 542 1860

Fax: 706 583 0210

Email: mmpratt@uga.edu

Seq primer: PolyTMix

High quality sequence start: 11

High quality sequence stop: 440

POLYA-No.

Location/Qualifiers

1..446

/organism="Sorghum propinquum"

/db.xref="taxon:132711"

/clone.lib="Rhizome2 (RH122)"

/note="Organ: Rhizomes; Vector: pBluescript II from Lambda

Zap II; Site_1: XhoI; Site_2: EcoRI; The library was made from poly-A RNA in the cloning vector lambda ZAP II. Clones to be sequenced were prepared by mass excision."

BASE COUNT 101 a 81 c 124 g 140 t

ORIGIN

Query Match

Best Local Similarity 59.0%; Pred. No. 1.1;

Matches 72; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

QY 345 CGCCAGAACTTCAGCGCCAGAGTTGGCGCGCCAGCGGTGTGTGGCGGTGT 404

DB 51 CGCGAGAGAGCTGACCGAGGAGGTGCGACGAGATGATCCGTGAGGCTGAGTCGATGG 110

QY 405 TCCTCCGCCAGCTTTCGGTAAAGGCTCTGCTGTGTCGATGCCATCGCGCGGTGCGGT 464

DB 111 CGACGCCAGATCAACTATGAAGAGTTTGTAAAGTTATGATGGCCAAAGTGAGGAGCGGT 170

QY 465 TC 466

DB 171 CC 172

RESULT 12

LOCUS

BF586874

DEFINITION

500 bp mRNA linear EST 12-DEC-2000

Floral-induced Meristem 1 (FM1) Sorghum

propinquum cDNA, mRNA sequence.

ACCESSION

BF586874

VERSION

BF586874.1 GI:11679198

KEYWORDS

EST.

SOURCE

Sorghum propinquum.

ORGANISM

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC

clade; Panicoideae; Andropogoneae; Sorghum.

REFERENCE

1 (bases 1 to 500)

AUTHORS

Cordonnier-Pratt,M.-M., Gingle,A., Sudman,M., Marsala,C. and Pratt

,L.H.

TITLE

An EST database from Sorghum: floral-induced meristems

JOURNAL

Unpublished (2000)

COMMENT

Contact: Cordonnier-Pratt MM

Laboratory for Genomics and Bioinformatics

The University of Georgia, Department of Plant Biology

Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA

Tel: 706 542 1860

Fax: 706 583 0210

Email: mmpratt@uga.edu

Seq primer: PolyTMix

High quality sequence start: 8

High quality sequence stop: 495

POLYA-No.

Location/Qualifiers

1..500

/organism="Sorghum propinquum"

/db.xref="taxon:132711"

/clone.lib="Floral-Induced Meristem 1 (FM1)"

/note="Organ: Floral-induced meristems; Vector:

pBluescript II from Lambda zap II; Site_1: XhoI; Site_2:

EcoRI; mature plants were placed in a growth chamber for

15 days with 16 hr darkness and 8 hr light (flowering is

induced by short-day conditions); 16 days after being

returned to the greenhouse under natural long days during

late April/early May, meristems were harvested The

library was made from poly-A RNA in the cloning vector

lambda ZAP II. Clones to be sequenced were prepared by

mass excision."

BASE COUNT 111 a 94 c 141 g 154 t

ORIGIN

LOCUS BG053014 531 bp mRNA linear EST 25-JAN-2001
DEFINITION RHIZ2_16_D04.g1_A003 Rhizome2 (RHIZ2) Sorghum propinquum cDNA, mRNA sequence.
ACCESSION BG053014
VERSION BG053014.1 GI:12508270
KEYWORDS EST.
SOURCE Sorghum propinquum.
ORGANISM Sorghum propinquum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE 1. (bases 1 to 531)
AUTHORS Cordonnier-Pratt, M.-M., Gingle, A., Paterson, A., Sudman, M. and Pratt, L.
TITLE An EST database from Sorghum: Sorghum propinquum rhizomes
JOURNAL Unpublished (2000)
COMMENT Contact: Cordonnier-Pratt MM
Laboratory for Genomics and Bioinformatics
The University of Georgia, Department of Plant Biology
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
Tel: 706 542 1860
Fax: 706 583 0210
Email: mmpratt@uga.edu
Seq primer: polyTmix
High quality sequence start: 42
High quality sequence stop: 527
POLYA-No. Location/Qualifiers
1. .531
/organism="Sorghum propinquum"
/db_xref="taxon:132711"
/clone_lib="Rhizome2 (RHIZ2)"
/notes="Organ: Rhizomes; Vector: pBluescript II from Lambda Zap II; Site_1: XhoI; Site_2: EcoRI; The library was made from poly-A RNA in the cloning vector lambda Zap II. Clones to be sequenced were prepared by mass excision."
BASE COUNT 122 a 101 c 152 g 154 t 2 others
ORIGIN
Query Match 5.5%; Score 42; DB 12; Length 531;
Best Local Similarity 59.0%; Pred. No. 1.2;
Matches 72; Conservative 0; Mismatches 50; Indels 0; Gaps 0;
QY 345 CGCCCGACAGTTCAGCGCCACGAGAGTGGCGCGCCACCGGTGGCGGTGGCGGTGT 404
Db 121 CGCGCGAAGCTGACCGACGAGGAGTGCAGAGATGATCCGTGAGGCTGACGTGATGG 180
QY 405 TGCCGGCCAGCTGTTCGGTAAAGCTCTGTGTCGTGTGTCGATGCCCATCGCGCGTGC 464
Db 181 CGACGCCAGATCAACTATGAAGATTGTTAAGTTATGATGTCGCCAAGTCAGGAGCGGT 240
QY 465 TC 466
Db 241 CC 242
RESULT 15
LOCUS BG713844 537 bp mRNA linear EST 08-MAY-2001
DEFINITION EM1_20_C08.g2_A002 Embryo 1 (EM1) Sorghum bicolor cDNA, mRNA sequence.
ACCESSION BG713844
VERSION BG713844.1 GI:14007794
KEYWORDS EST.
SOURCE sorghum.
ORGANISM Sorghum bicolor
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoideae; Andropogoneae; Sorghum.

REFERENCE 1 (bases 1 to 537)
AUTHORS Reid,S.P., Cordonnier-Pratt,M.-M., Gingle,A. and Pratt,L.H.
TITLE An EST database from Sorghum: developing embryos
JOURNAL Unpublished (2000)
COMMENT Contact: Cordonnier-Pratt MM
Laboratory for Genomics and Bioinformatics
The University of Georgia, Department of Plant Biology
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
Tel: 706 542 1860
Fax: 706 583 0210
Email: mmpratt@uga.edu
Sequences have been trimmed to exclude PolyA, vector and regions
below Phred quality 16. The threshold for highest quality sequence
is 20.
Seq primer: PolyTMix
High quality sequence start: 62
High quality sequence stop: 536
POLYA-No. Location/Qualifiers
1..537
/organism="Sorghum bicolor"
/db_xref="taxon:4558"
/clone_lib="Embryo 1 (EM1)"
/note="Organ: Embryos germinated for 24 hr; Vector:
pBluescript II from Lambda Zap II; Site_1: XhoI; Site_2:
EcoRI; The library was made from poly-A RNA in the cloning
vector lambda ZAP II. Clones to be sequenced were
prepared by mass excision."
BASE COUNT 118 a 98 c 152 g 169 t
ORIGIN
Query Match 5.5%; Score 42; DB 12; Length 537;
Best Local Similarity 59.0%; Pred. NO. 1.2;
Matches 72; Conservative 0; Mismatches 50; Indels 0; Gaps 0;
QY 345 CGCCGAGAACTTCAGCGCCAGGAGTTGGCGGCCACCGGTGGTGTGGCGGTGT 404
DB 92 CGCGGAGAACTTCAGCGCCAGGAGTTGGCGGCCACCGGTGGTGTGGCGGTGT 151
QY 405 TGCCGGCCAGCTGTTCGTTAAAGGCTCTGGTGTGTGTGTGTGTGTGTGTGTGT 464
DB 152 CGACGCCAGATCACTATGAGAGTTTCTTAAGGTTATGATGGCAAGTGAGGCGGT 211
QY 465 TC 466
DB 212 CC 213

Search completed: February 22, 2003, 17:06:30
Job time : 1557.18 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 22, 2003, 14:36:28 ; Search time 63.6684 Seconds
(without alignments)
3699.288 Million cell updates/sec

Title: US-09-677-374-5

Perfect score: 768
Sequence: 1 atgtcagttgaattctacaa.....aggtgattagcacccgaaaaa 768

Scoring table: IDENTITY_NUC

Gapop 10.0 ; Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 200000000
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

100% Processing. Minimum Match 0%
 Maximum Match 100%
 Listing first 45 summaries

```
Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/
2: /cgn2_6/ptodata/
3: /cgn2_6/ptodata/
4: /cgn2_6/ptodata/
5: /cgn2_6/ptodata/
6: /cgn2_6/ptodata/
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query | | DB | ID | Description |
|------------|-------|-------|--------|----|-------------------|-------------------|
| | | Match | Length | | | |
| 1 | 270.2 | 35.2 | 486 | 1 | US-08-048-164A-1 | Sequence 1, Appli |
| C 2 | 270.2 | 35.2 | 486 | 1 | US-08-048-164A-3 | Sequence 3, Appli |
| 3 | 270.2 | 35.2 | 486 | 1 | US-08-460-462-1 | Sequence 1, Appli |
| C 4 | 270.2 | 35.2 | 486 | 1 | US-08-460-462-3 | Sequence 3, Appli |
| 5 | 270.2 | 35.2 | 486 | 1 | US-08-460-457-1 | Sequence 1, Appli |
| C 6 | 270.2 | 35.2 | 486 | 1 | US-08-460-457-3 | Sequence 3, Appli |
| 7 | 270.2 | 35.2 | 486 | 1 | US-08-460-458-1 | Sequence 1, Appli |
| C 8 | 270.2 | 35.2 | 486 | 1 | US-08-460-458-3 | Sequence 3, Appli |
| 9 | 270.2 | 35.2 | 486 | 2 | US-08-460-455-1 | Sequence 1, Appli |
| C 10 | 270.2 | 35.2 | 486 | 2 | US-08-460-455-3 | Sequence 3, Appli |
| 11 | 270.2 | 35.2 | 486 | 2 | US-08-330-394A-1 | Sequence 1, Appli |
| C 12 | 270.2 | 35.2 | 486 | 2 | US-08-330-394A-3 | Sequence 3, Appli |
| 13 | 270.2 | 35.2 | 499 | 3 | US-09-008-636-5 | Sequence 5, Appli |
| 14 | 270.2 | 35.2 | 499 | 4 | US-09-008-632-5 | Sequence 5, Appli |
| 15 | 270.2 | 35.2 | 499 | 4 | US-09-325-274-5 | Sequence 5, Appli |
| 16 | 270.2 | 35.2 | 1146 | 4 | US-09-277-716-21 | Sequence 21, Appl |
| 17 | 270.2 | 35.2 | 1146 | 4 | US-09-609-161B-21 | Sequence 21, Appl |
| 18 | 70.2 | 9.1 | 1482 | 4 | US-09-198-956-9 | Sequence 9, Appli |
| 19 | 70.2 | 9.1 | 1482 | 4 | US-09-198-955A-11 | Sequence 11, Appl |
| 20 | 70.2 | 9.1 | 1482 | 4 | US-09-694-531-11 | Sequence 11, Appl |
| 21 | 70.2 | 9.1 | 1482 | 4 | US-09-670-141-9 | Sequence 9, Appli |
| C 22 | 40.8 | 5.2 | 168575 | 4 | US-09-426-290-1 | Sequence 1, Appli |
| 23 | 40.8 | 5.2 | 2645 | 4 | US-08-960-780-31 | Sequence 31, Appl |
| 24 | 40 | 5.2 | 2645 | 4 | US-09-073-898-31 | Sequence 31, Appl |
| 25 | 40 | 5.2 | 2845 | 4 | US-09-371-913A-1 | Sequence 1, Appli |
| 26 | 39.4 | 5.1 | 2004 | 4 | US-08-471-933-6 | Sequence 6, Appli |
| 27 | 39.4 | 5.1 | 2004 | 2 | US-08-471-044-6 | Sequence 6, Appli |

Query Match 35.2%; Score 270.2; DB 1; Length 486;
Best Local Similarity 98.9%; Pred. No. 4.6e-73;
Matches 272; Conservative 0; Mismatches 3; Indels 0

| | | | | | | | | | |
|---------|-----|--------------|---|------------|---|--------|---|------|---|
| Matches | 272 | Conservative | 0 | Mismatches | 3 | Indels | 0 | Gaps | 0 |
|---------|-----|--------------|---|------------|---|--------|---|------|---|

ALIGNMENTS

```

, RESULT 1
, US-08-048-164A-1
, ; Sequence 1, Application US/08048164A
, ; Patent No. 5496934
, ; GENERAL INFORMATION:
, ; APPLICANT: Shoseyov, Oded
, ; APPLICANT: Shlegl, Itai
, ; APPLICANT: Goldstein, Marc A.
, ; APPLICANT: Doi, Roy H.
, ; TITLE OF INVENTION: CELLULOSE BINDING DOMAIN
, ; NUMBER OF SEQUENCES: 21
, ; CORRESPONDENCE ADDRESS:
, ; ADDRESSEE: PENNIE & EDMONDS
, ; STREET: 1155 Avenue of the Americas
, ; CITY: New York
, ; STATE: New York
, ; COUNTRY: U.S.A.
, ; ZIP: 10036

```

```

1  ZIP: 10050
2  COMPUTER READABLE FORM:
3  MEDIUM TYPE: Floppy disk
4  COMPUTER: IBM PC compatible
5  OPERATING SYSTEM: PC-DOS/MS-DOS
6  SOFTWARE: PatentIn Release #1.0, Version #1.25
7  CURRENT APPLICATION DATA:
8  APPLICATION NUMBER: US/08/048,164A
9  FILING DATE: 14-APR-1993
10 CLASSIFICATION: 435
11 ATTORNEY/AGENT INFORMATION:
12 NAME: Mistrock, S. Leslie
13 REGISTRATION NUMBER: 18,872
14 REFERENCE/DOCKET NUMBER: 7809-003
15 TELECOMMUNICATION INFORMATION:
16 TELEPHONE: (212) 790-9090
17 TELEFAX: (212) 869-8864/9741
18 TELEX: 66141 PENNIE
19 INFORMATION FOR SEQ ID NO: 1:
20 SEQUENCE CHARACTERISTICS:
21 LENGTH: 486 base pairs
22 TYPE: nucleic acid
23 STRANDEDNESS: double
24 TOPOLOGY: unknown
25 MOLECULE TYPE: DNA
26 FEATURE:
27 NAME/KEY: CDS
28 LOCATION: 1..486
29 US-08-048-164A-1

```

Query Match 35.2%; Score 270.2; DB 1; Length 486;
Best Local Similarity 98.9%; Pred. No. 4.6e-73;
Matches 272; Conservative 0; Mismatches 3; Indels 0

QY 1 ATGTCAGTTGAATTTACAACTCTACAACTTACAAATCAGCACAAACAACTCAATTACACCAATA 60
|||||
Db 16 ATGTCAGTTGAATTTACAACTCTACAAATCAGCACAAACAACTCAATTACACCAATA 75
|||||
QY 61 ATCAAAATTTACTAACACATCTGACAGTGATTTAAATTTAAATGACGTAAAGCTTAGATAT 120
|||||
Db 76 ATCAAAATTTACTAACACATCTGACAGTGATTTAAATTTAAATGACGTAAAGCTTAGATAT 135
|||||
QY 121 TATTACACAAGTGATGGTACACAAGGACAAACTTTCTGGTGTGACCATCTGGTGCATTA 180
|||||
Db 136 TATTACACAAGTGATGGTACACAAGGACAAACTTTCTGGTGTGACCATCTGGTGCATTA 195
|||||
QY 181 TTAGGAAATAGCTATGTTGATACACTAGCAAGTGACAGCAAACTTCGTTAAAGAAACA 240
|||||
Db 196 TTAGGAAATAGCTATGTTGATACACTAGCAAGTGACAGCAAACTTCGTTAAAGAAACA 255
|||||
QY 241 GCAAGCCCAACATCAACCTATGATACATATCTGGA 275
|||||
Db 256 GCAAGCCCAACATCAACCTATGATACATATGTTGA 290
|||||

RESULT 2

US-08-164A-3/c
; Sequence 3, Application US/08048164A
; Patent No. 5496934

GENERAL INFORMATION:

; APPLICANT: Shoseyov, Oded
; APPLICANT: Shpiegl, Itai
; APPLICANT: Goldstein, Marc A.
; APPLICANT: Doi, Roy H.
; TITLE OF INVENTION: CELLULOSE BINDING DOMAIN
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PENNIE & EDMONDS
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036

COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/048.164A
; FILING DATE: 14-APR-1993
; CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

; NAME: Mistock, S. Leslie
; REGISTRATION NUMBER: 18,872
; REFERENCE/DOCKET NUMBER: 7809-003
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-8864/9741

TELEX: 66141 PENNIE

INFORMATION FOR SEQ ID NO: 3:

; SEQUENCE CHARACTERISTICS:
; LENGTH: 486 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA
US-08-048-164A-3

Query Match 35.2%; Score 270.2; DB 1; Length 486;

Best Local Similarity 98.9%; Pred. No. 4.6e-73;

Matches 272; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ATGTCAGTTGAATTTACAACTCTACAACTTACAAATCAGCACAAACAACTCAATTACACCAATA 60
|||||

Db 471 ATGTCAGTTGAATTTACAACTCTACAAATCAGCACAAACAACTCAATTACACCAATA 412
|||||

QY 61 ATCAAAATTTACTAACACATCTGACAGTGATTTAAATTTAAATGACGTAAAGCTTAGATAT 120
|||||
Db 411 ATCAAAATTTACTAACACATCTGACAGTGATTTAAATTTAAATGACGTAAAGCTTAGATAT 352
|||||
QY 121 TATTACACAAGTGATGGTACACAAGGACAAACTTTCTGGTGTGACCATCTGGTGCATTA 180
|||||
Db 351 TATTACACAAGTGATGGTACACAAGGACAAACTTTCTGGTGTGACCATCTGGTGCATTA 292
|||||
QY 181 TTAGGAAATAGCTATGTTGATACACTAGCAAGTGACAGCAAACTTCGTTAAAGAAACA 240
|||||
Db 291 TTAGGAAATAGCTATGTTGATACACTAGCAAGTGACAGCAAACTTCGTTAAAGAAACA 232
|||||
QY 241 GCAAGCCCAACATCAACCTATGATACATATCTGGA 275
|||||
Db 231 GCAAGCCCAACATCAACCTATGATACATATGTTGA 197
|||||

RESULT 3

US-08-460-462-1
; Sequence 1, Application US/08460462
; Patent No. 5670623

GENERAL INFORMATION:

; APPLICANT: Shoseyov, Oded
; APPLICANT: Shpiegl, Itai
; APPLICANT: Goldstein, Marc A.
; APPLICANT: Doi, Roy H.
; TITLE OF INVENTION: METHODS OF USE OF CELLULOSE BINDING DOMAIN PROTEINS
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:

; ADDRESSEE: PENNIE & EDMONDS
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036

COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/460.462
; FILING DATE: Concurrently herewith
; CLASSIFICATION: 435

PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/048.164
; FILING DATE: 14-APR-1993

ATTORNEY/AGENT INFORMATION:

; NAME: Mistock, S. Leslie
; REGISTRATION NUMBER: 18,872
; REFERENCE/DOCKET NUMBER: 7809-006
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-8864/9741

TELEX: 66141 PENNIE

INFORMATION FOR SEQ ID NO: 1:

; SEQUENCE CHARACTERISTICS:
; LENGTH: 486 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..486
US-08-460-462-1

Query Match 35.2%; Score 270.2; DB 1; Length 486;

Best Local Similarity 98.9%; Pred. No. 4.6e-73;

Matches 272; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ATGTCAGTTGAATTTACAACTCTACAACTTACAAATCAGCACAAACAACTCAATTACACCAATA 60
|||||

| | | | |
|-----|----|---|-----|
| 471 | Db | ATGTCAGTTGAATTTTACAACCTCTAACAAATCAGCACAACAACATCAATTTACACCAATA | 412 |
| 61 | Qy | ATCAAAATTACTAACACATCTGACAGTGAATTTAAATTTAAATGACGTAAAGTTAGATAT | 120 |
| 411 | Db | ATCAAAATTACTAACACATCTGACAGTGAATTTAAATTTAAATCAGCTAAAGTTACATAT | 352 |
| 121 | Qy | TATTACACAAGTGATGGTACACAAGGACAAACCTTTCTGGTGTGACCATGCTGGTGCATTA | 180 |
| 351 | Db | TATTACACAAGTGATGGTACACAAGGACAAACCTTTCTGGTGTGACCATGCTGGTGCATTA | 292 |
| 181 | Qy | TTAGGAAATAGCTATGTTGTATACACTAGCAAAAGTCAGACAAACCTTCGTTAAAGAAACA | 240 |
| 291 | Db | TTAGGAAATAGCTATGTTGTATACACTAGCAAAAGTCAGACAAACCTTCGTTAAAGAAACA | 332 |
| 241 | Qy | GCAAGCCCAACATCAACCTTATGATACATATCTCGA | 275 |
| 231 | Db | GCAAGCCCAACATCAACCTTATGATACATATGTTGA | 197 |

RESULT 5

```

US-08-460-457-1
; Sequence 1, Application US/08460457
; Patent No. 5719044
; GENERAL INFORMATION:
; APPLICANT: Shoseyov, Oded
; APPLICANT: Shpiegl, Itai
; APPLICANT: Goldstein, Marc A.
; APPLICANT: Doi, Roy H.
; TITLE OF INVENTION: CELLULOSE BINDING DOMAIN FUSION PROTEINS
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PENNIE & EDMONDS
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/460,457
; FILING DATE: concurrently herewith
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/048,164
; FILING DATE: 14-APR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Mistock, S. Leslie
; REGISTRATION NUMBER: 18,872
; REFERENCE/DOCKET NUMBER: 7809-008
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-8864/9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 486 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..486
US-08-460-457-1

```

Query Match 35.2%; Score 270.2; DB 1; Length 486;
 Best Local Similarity 98.9%; Pred. No. 4.6e-73;
 Matches 272; Conservative 0; Mismatches 3; Indels 0

Best Local Similarity 98.9%; Pred. No. 4.6e-73;
Matches 272; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

| | | | |
|----|-----|--|-----|
| QY | 1 | ATGTCAGTTGAAATCTACAACTCTACAAATCAGCAACAACAACTCAATTTACACCAATA | 60 |
| Db | 16 | ATGTCAGTTGAAATTTTACAACTCTAACAATCAGCAACAACAACTCAATTTACACCAATA | 75 |
| QY | 61 | ATCAAAATTAATTAACACATCTGACAGTGATTTAAATTTAAATGACCTAAAAGTTAGATAT | 120 |
| Db | 76 | ATCAAAATTAATTAACACATCTGACAGTGATTTAAATTTAAATGACCTAAAAGTTAGATAT | 135 |
| QY | 121 | TATTACAAAGTGGTGTACACAAGGACAAACCTTTCTGGTGTGACCATGCTGGTGCATTTA | 180 |
| Db | 136 | TATTACAAAGTGGTGTACACAAGGACAAACCTTTCTGGTGTGACCATGCTGGTGCATTTA | 195 |
| QY | 181 | TTAGGAAATAGCTATGTTGATTAACACTAGCAAAAGTCACAGCAAACTTCGTTAAAGAAACA | 240 |
| Db | 196 | TTAGGAAATAGCTATGTTGATTAACACTAGCAAAAGTCACAGCAAACTTCGTTAAAGAAACA | 255 |
| QY | 241 | GCAAGCCCAACATCAACCTATGATACATATCTCGA | 275 |
| Db | 256 | GCAAGCCCAACATCAACCTATGATACATATGTTGA | 290 |

RESULT 6

US-08-460-457-3/C
; Sequence 3, Application US/08460457
; Patent No. 5719044
; GENERAL INFORMATION:
; APPLICANT: Shoseyov, Oded
; APPLICANT: Shpiegl, Itai
; APPLICANT: Goldstein, Marc A.
; APPLICANT: Doi, Roy H.
; TITLE OF INVENTION: CELLULOSE BINDING DOMAIN FUSION PROTEINS
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PENNIE & EDMONDS
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/460,457
; FILING DATE: concurrently herewith
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/048,164
; FILING DATE: 14-APR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Mistrock, S. Leslie
; REGISTRATION NUMBER: 18,872
; REFERENCE/DOCKET NUMBER: 7809-008
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-8864/9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 486 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA
; US-08-460-457-3

[illegible]

RESULT 7

```

US-08-460-458-1
Sequence 1, Application US/08460458
Patent No. 5738984
GENERAL INFORMATION:
APPLICANT: Shoseyov, Oded
TITLE OF INVENTION: KITS AND METHODS OF DETECTION
NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESS:
ADDRESSEE: PENNIE & EDMONDS
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/460,458
FILING DATE: concurrently herewith
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/048,164
FILING DATE: 14-APR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Masrock, S. Leslie
REGISTRATION NUMBER: 18,972
REFERENCE/DOCKET NUMBER: 7809-007
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-8864/9741
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 486 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: unknown
MOLECULE TYPE: DNA
FEATURE:
NAME/KEY: CDS
LOCATION: 1..486
US-08-460-458-1

```

Db 16 ATGTCAGTTGAATTTTACAACTCTAACAAATCAGACAAACAACTCAATTACACCAATA 75
QY 61 ATCAAAATTAACACATCTGACAGTGATTTAAATTTAAATGACGTAAAGTTAGATAT 120
Db 76 ATCAAAATTAACACATCTGACAGTGATTTAAATTTAAATGACGTAAAGTTAGATAT 135
QY 121 TATTACACAAGTGATGTTGATACACAAGGACAACTTTCTGGTGTGACCATGCTGTGCAATTA 180
Db 136 TATTACACAAGTGATGTTGATACACAAGGACAACTTTCTGGTGTGACCATGCTGTGCAATTA 195
QY 181 TTAGGAAATAGCTATCTGATACACTAGCAAAAGTGACGCAAACTTCGTTAAAGAAACA 240
Db 196 TTAGGAAATAGCTATCTGATACACTAGCAAAAGTGACGCAAACTTCGTTAAAGAAACA 255
QY 241 GCAAGCCCAACATCAACCTATGATACATATCTGGA 275
Db 256 GCAAGCCCAACATCAACCTATGATACATATGTTGA 290

RESULT 8
US-08-460-458-3/c
; Sequence 3, Application US/08460458
; Patent No. 5738984
; GENERAL INFORMATION:
; APPLICANT: Shoseyov, Oded
; TITLE OF INVENTION: KITS AND METHODS OF DETECTION USING CELLULOSE BINDING DOMAIN
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PENNIE & EDMONDS
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/460,458
; FILING DATE: concurrently herewith
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/048,164
; FILING DATE: 14-APR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Misrock, S. Leslie
; REGISTRATION NUMBER: 18,872
; REFERENCE/DOCKET NUMBER: 7809-007
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-8864/9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 486 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA
US-08-460-458-3

Query Match 35.2%; Score 270.2; DB 1; Length 486;
Best Local Similarity 98.9%; Pred. No. 4.6e-73;
Matches 272; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 ATGTCAGTTGAATTTTACAACTCTAACAAATCAGCACAACAACTCAATTACACCAATA 60
Db 471 ATGTCAGTTGAATTTTACAACTCTAACAAATCAGCACAACAACTCAATTACACCAATA 412
QY 61 ATCAAAATTAACACATCTGACAGTGATTTAAATTTAAATGACGTAAAGTTAGATAT 120

Db 411 ATCAAAATTAACACATCTGACAGTGATTTAAATTTAAATGACGTAAAGTTAGATAT 352
QY 121 TATTACACAAGTGATGTTGATACACAAGGACAACTTTCTGGTGTGACCATGCTGTGCAATTA 180
Db 351 TATTACACAAGTGATGTTGATACACAAGGACAACTTTCTGGTGTGACCATGCTGTGCAATTA 292
QY 181 TTAGGAAATAGCTATCTGATACACTAGCAAAAGTGACGCAAACTTCGTTAAAGAAACA 240
Db 291 TTAGGAAATAGCTATCTGATACACTAGCAAAAGTGACGCAAACTTCGTTAAAGAAACA 232
QY 241 GCAAGCCCAACATCAACCTATGATACATATCTGGA 275
Db 231 GCAAGCCCAACATCAACCTATGATACATATGTTGA 197

RESULT 9
US-08-460-455-1
; Sequence 1, Application US/08460455
; Patent No. 5837814
; GENERAL INFORMATION:
; APPLICANT: Shoseyov, Oded
; APPLICANT: Shpiegl, Itai
; APPLICANT: Goldstein, Marc A.
; APPLICANT: Doi, Roy H.
; TITLE OF INVENTION: CELLULOSE BINDING DOMAIN PROTEINS
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PENNIE & EDMONDS
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/460,455
; FILING DATE: concurrently herewith
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/048,164
; FILING DATE: 14-APR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Misrock, S. Leslie
; REGISTRATION NUMBER: 18,872
; REFERENCE/DOCKET NUMBER: 7809-009
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-8864/9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 486 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..486
US-08-460-455-1

Query Match 35.2%; Score 270.2; DB 2; Length 486;
Best Local Similarity 98.9%; Pred. No. 4.6e-73;
Matches 272; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 ATGTCAGTTGAATTTTACAACTCTAACAAATCAGCACAACAACTCAATTACACCAATA 60
Db 16 ATGTCAGTTGAATTTTACAACTCTAACAAATCAGCACAACAACTCAATTACACCAATA 75

QY 61 ATCAAAATTTACTAACACATCTGACAGTGTATTTAAATTTAAATGACGTAAGTAAAGTTAGATAT 120
|||||
Db 76 ATCAAAATTTACTAACACATCTGACAGTGTATTTAAATTTAAATGACGTAAGTAAAGTTAGATAT 135
|||||
QY 121 TATTACAAAGTGATGGTACACAAGGACAAACTTTCTGGTGTGACCATGCTGGTGCATTA 180
|||||
Db 136 TATTACAAAGTGATGGTACACAAGGACAAACTTTCTGGTGTGACCATGCTGGTGCATTA 195
|||||
QY 181 TTAGGAATAGCTATGTTGATTAACACTAGCAAAAGTGACAGCAAACTTCGTTAAAGAAACA 240
|||||
Db 196 TTAGGAATAGCTATGTTGATTAACACTAGCAAAAGTGACAGCAAACTTCGTTAAAGAAACA 255
|||||
QY 241 GCAAGCCCAACATCAACCTATGATACATATCTGGA 275
|||||
Db 256 GCAAGCCCAACATCAACCTATGATACATATGTTGA 290
|||||

RESULT 10

US-08-460-455-3/c
; Sequence 3, Application US/08460455
; Patent No. 5837814
; GENERAL INFORMATION:
; APPLICANT: Shoseyov, Oded
; APPLICANT: Yosef, Karmey
; APPLICANT: Shpiegl, Itai
; APPLICANT: Goldstein, Marc A.
; APPLICANT: Doi, Roy H.
; TITLE OF INVENTION: CELLULOSE BINDING DOMAIN PROTEINS
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PENNIE & EDMONDS
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/460.455
; FILING DATE: Concurrently herewith
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/048.164
; FILING DATE: 14-APR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Mistock, S. Leslie
; REGISTRATION NUMBER: 18,872
; REFERENCE/DOCKET NUMBER: 7809-009
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-8864/9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 486 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA
US-08-460-455-3

Query Match 35.2%; Score 270.2; DB 2; Length 486;
Best Local Similarity 98.9%; Pred. No. 4.6e-73;
Matches 272; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 ATGTCAGTTGAATTTACAACTCTAACAAATCAGCACAAACAACTCAATTACACCAATA 60
|||||
Db 471 ATGTCAGTTGAATTTTACAACTCTAACAAATCAGCACAAACAACTCAATTACACCAATA 412
|||||
QY 61 ATCAAAATTTACTAACACATCTGACAGTGTATTTAAATTTAAATGACGTAAGTAAAGTTAGATAT 120
|||||

Db 411 ATCAAAATTTACTAACACATCTGACAGTGTATTTAAATTTAAATGACGTAAGTAAAGTTAGATAT 352
|||||
QY 121 TATTACAAAGTGATGGTACACAAGGACAAACTTTCTGGTGTGACCATGCTGGTGCATTA 180
|||||
Db 351 TATTACAAAGTGATGGTACACAAGGACAAACTTTCTGGTGTGACCATGCTGGTGCATTA 292
|||||
QY 181 TTAGGAATAGCTATGTTGATTAACACTAGCAAAAGTGACAGCAAACTTCGTTAAAGAAACA 240
|||||
Db 291 TTAGGAATAGCTATGTTGATTAACACTAGCAAAAGTGACAGCAAACTTCGTTAAAGAAACA 232
|||||
QY 241 GCAAGCCCAACATCAACCTATGATACATATCTGGA 275
|||||
Db 231 GCAAGCCCAACATCAACCTATGATACATATGTTGA 197
|||||

RESULT 11

US-08-330-394A-1
; Sequence 1, Application US/08330394A
; Patent No. 5856201
; GENERAL INFORMATION:
; APPLICANT: Shoseyov, Oded
; APPLICANT: Yosef, Karmey
; APPLICANT: Shpiegl, Itai
; APPLICANT: Goldstein, Marc A.
; APPLICANT: Doi, Roy H.
; TITLE OF INVENTION: METHODS OF DETECTION USING THE
; TITLE OF INVENTION: CELLULOSE BINDING DOMAIN
; NUMBER OF SEQUENCES: 33
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PENNIE & EDMONDS
; STREET: 1155 AVENUE OF THE AMERICAS
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/330.394A
; FILING DATE: 27-OCT-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: CIP OF PCT/US94/04132
; FILING DATE: 14-APR-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: MISROCK, S. LESLIE
; REFERENCE/DOCKET NUMBER: 7809-005
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-8864
; TELEX: 66441 PENNIE
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 486 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..486
US-08-330-394A-1

Query Match 35.2%; Score 270.2; DB 2; Length 486;
Best Local Similarity 98.9%; Pred. No. 4.6e-73;
Matches 272; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 ATGTCAGTTGAATTTTACAACTCTAACAAATCAGCACAAACAACTCAATTACACCAATA 60
|||||
Db 16 ATGTCAGTTGAATTTTACAACTCTAACAAATCAGCACAAACAACTCAATTACACCAATA 75
|||||

QY 61 ATCAAAATTACTACACATCTGACAGTGATTTAAATTTAAATGACGTAAAGTTAGATAT 120
|||||
Db 76 ATCAAAATTACTACACATCTGACAGTGATTTAAATTTAAATGACGTAAAGTTAGATAT 135
|||||
QY 121 TATTACACAAGTGATGTACACAAGGACAAACTTTCTGGTGTGACCATGCTGGTGCATTA 180
|||||
Db 136 TATTACACAAGTGATGTACACAAGGACAAACTTTCTGGTGTGACCATGCTGGTGCATTA 195
|||||
QY 181 TTAGGAATAGCTATGTTGATAACACTAGCAAAAGTGACGCAAACTTCGTTAAAGAAACA 240
|||||
Db 196 TTAGGAATAGCTATGTTGATAACACTAGCAAAAGTGACGCAAACTTCGTTAAAGAAACA 255
|||||
QY 241 GCAAGCCCAACATCAACCTATGATACATATCTGGA 275
|||||
Db 256 GCAAGCCCAACATCAACCTATGATACATATGTTGA 290
|||||

RESULT 12

US-08-330-394A-3/C

; Sequence 3, Application US/08330394A

; Patent No. 5856201

; GENERAL INFORMATION:

; APPLICANT: Shoseyov, Oded

; APPLICANT: Yosef, Karmey

; APPLICANT: Shpiegl, Itai

; APPLICANT: Goldstein, Marc A.

; APPLICANT: Dol, Roy H.

; TITLE OF INVENTION: METHODS OF DETECTION USING THE

; TITLE OF INVENTION: CELLULOSE BINDING DOMAIN

; NUMBER OF SEQUENCES: 33

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: PENNIE & EDMONDS

; STREET: 1155 AVENUE OF THE AMERICAS

; CITY: NEW YORK

; STATE: NEW YORK

; COUNTRY: U.S.A.

; ZIP: 10036

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/330,394A

; FILING DATE: 27-OCT-1994

; PRIORITY DATA:

; APPLICATION NUMBER: CIP OF PCT/US94/04132

; FILING DATE: 14-APR-1994

; ATTORNEY/AGENT INFORMATION:

; NAME: MISROCK, S. LESLIE

; REFERENCE/DOCKET NUMBER: 7809-005

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (212) 790-9090

; TELEFAX: (212) 869-8864

; TELEX: 66441 PENNIE

; INFORMATION FOR SEQ ID NO: 3:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 486 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: double

; TOPOLOGY: unknown

; MOLECULE TYPE: DNA

US-08-330-394A-3

Query Match 35.2%; Score 270.2; DB 2; Length 486;

Best Local Similarity 98.9%; Pred. No. 4.6e-73;

Matches 272; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ATGTCACTTGAATTTCTACAACCTCTAACAATGACGCAAACTCAATTTACACCAATA 60
|||||
Db 471 ATGTCACTTGAATTTCTACAACCTCTAACAATGACGCAAACTCAATTTACACCAATA 412
|||||

QY 61 ATCAAAATTACTACACATCTGACAGTGATTTAAATTTAAATGACGTAAAGTTAGATAT 120
|||||
Db 411 ATCAAAATTACTACACATCTGACAGTGATTTAAATTTAAATGACGTAAAGTTAGATAT 352
|||||
QY 121 TATTACACAAGTGATGTACACAAGGACAAACTTTCTGGTGTGACCATGCTGGTGCATTA 180
|||||
Db 351 TATTACACAAGTGATGTACACAAGGACAAACTTTCTGGTGTGACCATGCTGGTGCATTA 292
|||||
QY 181 TTAGGAATAGCTATGTTGATAACACTAGCAAAAGTGACGCAAACTTCGTTAAAGAAACA 240
|||||
Db 291 TTAGGAATAGCTATGTTGATAACACTAGCAAAAGTGACGCAAACTTCGTTAAAGAAACA 232
|||||
QY 241 GCAAGCCCAACATCAACCTATGATACATATCTGGA 275
|||||
Db 231 GCAAGCCCAACATCAACCTATGATACATATGTTGA 197
|||||

RESULT 13

US-09-006-636-5

; Sequence 5, Application US/09006636

; Patent No. 6005092

; GENERAL INFORMATION:

; APPLICANT: Shoseyov, Oded

; APPLICANT: Shani, Ziv

; TITLE OF INVENTION: ARABIDOPSIS THALIANA ENDO-1,4-BETA-

; TITLE OF INVENTION: GLUCANASE GENE, PROMOTER AND PROTEIN

; NUMBER OF SEQUENCES: 21

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Pennie & Edmonds LLP

; STREET: 1155 Avenue of the Americas

; CITY: New York

; STATE: NY

; COUNTRY: USA

; ZIP: 10036-2711

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette

; COMPUTER: IBM Compatible

; OPERATING SYSTEM: DOS

; SOFTWARE: FastSeq Version 2.0

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/006,636

; FILING DATE:

; CLASSIFICATION:

; ATTORNEY/AGENT INFORMATION:

; NAME: Baldwin, Geraldine F.

; REGISTRATION NUMBER: 31,232

; REFERENCE/DOCKET NUMBER: 7809-019

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (212) 790-9090

; TELEFAX: (212) 869-8864

; TELEX: 66141 PENNIE

; INFORMATION FOR SEQ ID NO: 5:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 499 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: double

; TOPOLOGY: linear

; MOLECULE TYPE: DNA

US-09-006-636-5

Query Match 35.2%; Score 270.2; DB 3; Length 499;

Best Local Similarity 98.9%; Pred. No. 4.6e-73;

Matches 272; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ATGTCACTTGAATTTCTACAACCTCTAACAATGACGCAAACTCAATTTACACCAATA 60
|||||
Db 21 ATGTCACTTGAATTTCTACAACCTCTAACAATGACGCAAACTCAATTTACACCAATA 80
|||||
QY 61 ATCAAAATTACTACACATCTGACAGTGATTTAAATTTAAATGACGTAAAGTTAGATAT 120
|||||
Db 81 ATCAAAATTACTACACATCTGACAGTGATTTAAATTTAAATGACGTAAAGTTAGATAT 140
|||||
QY 121 TATTACACAAGTGATGTACACAAGGACAAACTTTCTGGTGTGACCATGCTGGTGCATTA 180
|||||

Db 141 TATTACAAAGTGATGGTACAAAGGACAAACTTTCTGGTGTGACCATCTGCTGGTGCATT 200
QY 181 TTAGGAATAGCTATGTTGATAAAGTACAAAGTACAGCAAACTTCGTTAAAGAAACA 240
Db 201 TTAGGAATAGCTATGTTGATAAAGTACAAAGTACAGCAAACTTCGTTAAAGAAACA 260
QY 241 GCAAGCCCAACATCAACCTATGATACATATCTGGA 275
Db 261 GCAAGCCCAACATCAACCTATGATACATATGTTGA 295

RESULT 14
US-09-006-632-5
; Sequence 5, Application US/09006632
; Patent No. 6184440
; GENERAL INFORMATION:
; APPLICANT: Shoseyov, Oded
; APPLICANT: Shani, Ziv
; TITLE OF INVENTION: TRANSGENIC PLANTS OF ALTERED
; TITLE OF INVENTION: MORPHOLOGY
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/006,632
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Baldwin, Geraldine F.
; REGISTRATION NUMBER: 31,232
; REFERENCE/DOCKET NUMBER: 7809-018
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-8864
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 499 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA
US-09-006-632-5

Query Match 35.2%; Score 270.2; DB 4; Length 499;
Best Local Similarity 98.9%; Pred. No. 4.6e-73;
Matches 272; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ATGTCAGTTGAATTTACAACTCTAACAAATCAGCACAAACAACTCAATTACACCAATA 50
Db 21 ATGTCAGTTGAATTTACAACTCTAACAAATCAGCACAAACAACTCAATTACACCAATA 80
QY 61 ATCAAAATTTACTTAACACATCTGACAGTGAATTTAAATTTAAATGAGCTAAAGTTAGATAT 120
Db 81 ATCAAAATTTACTTAACACATCTGACAGTGAATTTAAATTTAAATGAGCTAAAGTTAGATAT 140
QY 121 TATTACAAAGTGATGGTACAAAGGACAAACTTTCTGGTGTGACCATCTGCTGGTGCATT 180
Db 141 TATTACAAAGTGATGGTACAAAGGACAAACTTTCTGGTGTGACCATCTGCTGGTGCATT 200
QY 181 TTAGGAATAGCTATGTTGATAAAGTACAAAGTACAGCAAACTTCGTTAAAGAAACA 240

Db 201 TTAGGAATAGCTATGTTGATAAAGTACAAAGTACAGCAAACTTCGTTAAAGAAACA 260
QY 241 GCAAGCCCAACATCAACCTATGATACATATCTGGA 275
Db 261 GCAAGCCCAACATCAACCTATGATACATATGTTGA 295

RESULT 15
US-09-325-274-5
; Sequence 5, Application US/09325274
; Patent No. 6323023
; GENERAL INFORMATION:
; APPLICANT: Shoseyov, Oded
; APPLICANT: Shani, Ziv
; TITLE OF INVENTION: ARABIDOPSIS THALIANA ENDO-1,4-BETA-
; TITLE OF INVENTION: GLUCANASE GENE, PROMOTER AND PROTEIN
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/325,274
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/006,636
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Baldwin, Geraldine F.
; REGISTRATION NUMBER: 31,232
; REFERENCE/DOCKET NUMBER: 7809-019
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-8864
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 499 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA
US-09-325-274-5

Query Match 35.2%; Score 270.2; DB 4; Length 499;
Best Local Similarity 98.9%; Pred. No. 4.6e-73;
Matches 272; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ATGTCAGTTGAATTTACAACTCTAACAAATCAGCACAAACAACTCAATTACACCAATA 60
Db 21 ATGTCAGTTGAATTTACAACTCTAACAAATCAGCACAAACAACTCAATTACACCAATA 80
QY 61 ATCAAAATTTACTTAACACATCTGACAGTGAATTTAAATTTAAATGAGCTAAAGTTAGATAT 120
Db 81 ATCAAAATTTACTTAACACATCTGACAGTGAATTTAAATTTAAATGAGCTAAAGTTAGATAT 140
QY 121 TATTACAAAGTGATGGTACAAAGGACAAACTTTCTGGTGTGACCATCTGCTGGTGCATT 180
Db 141 TATTACAAAGTGATGGTACAAAGGACAAACTTTCTGGTGTGACCATCTGCTGGTGCATT 200
QY 181 TTAGGAATAGCTATGTTGATAAAGTACAAAGTACAGCAAACTTCGTTAAAGAAACA 240
Db 201 TTAGGAATAGCTATGTTGATAAAGTACAGCAAACTTCGTTAAAGAAACA 260

Search completed: February 22, 2003, 18:48:24
Job time : 69.6684 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.4_p5-4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 22, 2003, 16:18:29 ; Search time 74.7219 Seconds
(without alignments)
5772.590 Million cell updates/sec

Title: US-09-677-374-5
Perfect score: 768
Sequence: 1 atgtcagtgaattctacaa.....agggtattagaccgaaaaa 768

Scoring table:
IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 442118 seqs, 280819700 residues

Total number of hits satisfying chosen parameters: 884236

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

- Database : Published Applications_NA.*
- 1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq.*
 - 2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq.*
 - 3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq.*
 - 4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq.*
 - 5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq.*
 - 6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq.*
 - 7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq.*
 - 8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq.*
 - 9: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq.*
 - 10: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq.*
 - 11: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq.*
 - 12: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq.*
 - 13: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq.*
 - 14: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match % | Length | DB ID | Description |
|------------|-------|---------------|--------|-------|---------------------|
| 1 | 70.2 | 9.1 | 1482 | 12 | US-10-072-152-11 |
| 2 | 44 | 5.7 | 3509 | 10 | US-09-919-060-4 |
| 3 | 44 | 5.7 | 3509 | 10 | US-09-919-060-6 |
| c | 41.6 | 5.4 | 3195 | 12 | US-10-044-090-394 |
| 5 | 40 | 5.2 | 2645 | 10 | US-09-850-351A-31 |
| 6 | 40 | 5.2 | 2645 | 10 | US-09-967-805-1 |
| 7 | 38.8 | 5.1 | 289 | 9 | US-09-854-133-533 |
| 8 | 38.8 | 5.1 | 289 | 10 | US-09-738-973-533 |
| c | 38.4 | 5.0 | 500 | 10 | US-09-864-761-11900 |
| c | 37.4 | 4.9 | 480 | 10 | US-09-960-352-5301 |
| 11 | 36 | 4.7 | 7791 | 10 | US-09-764-877-3700 |
| c | 35.6 | 4.6 | 2481 | 10 | US-09-729-674-141 |
| 13 | 35.2 | 4.6 | 22484 | 10 | US-09-875-114-2 |
| 14 | 35.2 | 4.6 | 22484 | 10 | US-09-880-107-3341 |
| 15 | 35.2 | 4.6 | 335913 | 9 | US-09-754-853A-2 |
| 16 | 35.2 | 4.6 | 335913 | 9 | US-09-754-853A-3 |
| 17 | 35 | 4.6 | 70768 | 9 | US-10-135-322-13 |
| c | 34.8 | 4.5 | 413 | 10 | US-09-960-352-2919 |
| 19 | 34.8 | 4.5 | 426 | 9 | US-10-125-815-6 |

Query Match 9.1%; Score 70.2; DB 12; Length 1482;
Best Local Similarity 63.2%; Pred. No. 2.3e-10;

| | | | | | | | |
|---|----|------|-----|--------|----|---------------------|-------------------|
| c | 20 | 34.8 | 4.5 | 1138 | 9 | US-09-938-842A-4079 | Sequence 4079, Ap |
| c | 21 | 34.8 | 4.5 | 2002 | 10 | US-09-887-576-298 | Sequence 298, App |
| c | 22 | 34.8 | 4.5 | 2004 | 10 | US-09-887-576-299 | Sequence 299, App |
| c | 23 | 34.8 | 4.5 | 640681 | 10 | US-09-790-988-1 | Sequence 1, Appli |
| c | 24 | 34.6 | 4.5 | 4059 | 9 | US-09-784-554B-1 | Sequence 1, Appli |
| c | 25 | 34.4 | 4.5 | 1720 | 10 | US-09-887-576-583 | Sequence 583, App |
| c | 26 | 34.4 | 4.5 | 3694 | 10 | US-09-880-107-1816 | Sequence 1816, Ap |
| c | 27 | 34.4 | 4.5 | 7311 | 12 | US-10-044-090-435 | Sequence 455, App |
| c | 28 | 34.4 | 4.5 | 9172 | 10 | US-09-070-927A-65 | Sequence 65, Appl |
| c | 29 | 34.4 | 4.5 | 13737 | 9 | US-10-074-279-10 | Sequence 10, Appl |
| c | 30 | 34.2 | 4.5 | 1207 | 9 | US-09-938-842A-5018 | Sequence 5018, Ap |
| c | 31 | 33.8 | 4.4 | 810 | 9 | US-09-738-626-1633 | Sequence 1633, Ap |
| c | 32 | 33.8 | 4.4 | 45845 | 10 | US-09-927-091-6 | Sequence 6, Appli |
| c | 33 | 33.8 | 4.4 | 180557 | 12 | US-10-003-808-6 | Sequence 6, Appli |
| c | 34 | 33.8 | 4.4 | 180557 | 12 | US-10-003-808-9 | Sequence 9, Appli |
| c | 35 | 33.6 | 4.4 | 1164 | 9 | US-09-738-626-2633 | Sequence 2633, Ap |
| c | 36 | 33.6 | 4.4 | 3111 | 10 | US-09-815-242-4518 | Sequence 4518, Ap |
| c | 37 | 33.6 | 4.4 | 3198 | 10 | US-09-815-242-8519 | Sequence 8519, Ap |
| c | 38 | 33.6 | 4.4 | 9767 | 10 | US-09-070-927A-130 | Sequence 130, App |
| c | 39 | 33.4 | 4.3 | 2000 | 9 | US-09-938-842A-3270 | Sequence 3270, Ap |
| c | 40 | 33.2 | 4.3 | 416 | 10 | US-09-960-352-4584 | Sequence 4584, Ap |
| c | 41 | 33.2 | 4.3 | 1047 | 10 | US-09-780-996-6 | Sequence 6, Appli |
| c | 42 | 33.2 | 4.3 | 1278 | 9 | US-10-086-738A-1 | Sequence 1, Appli |
| c | 43 | 33.2 | 4.3 | 1792 | 10 | US-09-822-830A-420 | Sequence 420, App |
| c | 44 | 33.2 | 4.3 | 2187 | 9 | US-09-870-759-81 | Sequence 81, Appl |
| c | 45 | 33.2 | 4.3 | 9172 | 10 | US-09-070-927A-65 | Sequence 65, Appl |

ALIGNMENTS

RESULT 1
US-10-072-152-11
; Sequence 11, Application US/10072152
; Patent No. US20020142438A1
; GENERAL INFORMATION:
; APPLICANT: Andersen, Lene N.
; APPLICANT: Schuelein, Martin
; APPLICANT: Lange, Niels E.
; APPLICANT: Bjornvad, Mads E.
; APPLICANT: Moller, Soren
; APPLICANT: Glad, Sanne O. S.
; APPLICANT: Kauppinen, Markus S.
; APPLICANT: Schnorr, Kirk
; APPLICANT: Kongsbak, Lars
; TITLE OF INVENTION: No. US20020142438A1e1 Pectate Lyases
; FILE REFERENCE: 5378.200-US
; CURRENT APPLICATION NUMBER: US/10/072,152
; CURRENT FILING DATE: 2002-02-07
; PRIOR APPLICATION NUMBER: US/09/198,955
; PRIOR FILING DATE: 1998-11-24
; PRIOR APPLICATION NUMBER: 1343/97
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 1344/97
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 60/067,249
; PRIOR FILING DATE: 1997-12-02
; PRIOR APPLICATION NUMBER: 60/067,240
; PRIOR FILING DATE: 1997-12-02
; PRIOR APPLICATION NUMBER: 09/073,684
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: 09/184,217
; PRIOR FILING DATE: 1998-11-02
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 1482
; TYPE: DNA
; ORGANISM: Clostridium thermocellum
US-10-072-152-11

```

; LENGTH: 3509
; TYPE: DNA
; ORGANISM: Canis familiaris
US-09-919-060-6

Query Match      5.7%; Score 44; DB 10; Length 3509;
Best Local Similarity 56.1%; Pred. No. 0.014;
Matches 83; Conservative 0; Mismatches 65; Indels 0; Gaps 0;

QY 46 TCAATTACACCAATTAATCAAAATTACTTAACACATCTGACAGTGATTTAAATTTAAATGAC 105
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1430 TGAATTTAGCTTTATATTATTATGAACAAAGCGAAGATGTTGGATATTTAAATGCT 1371
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 106 GTAAAGTTAGATATATTATACACAGTAGTGTCACACAGGACAAACTTCTCGGTGTCGAC 165
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1370 GTCCAGGATGACAAATGCTCGAAGTTTTTTCGACACTATCGGGATTTCTAGTGATC 1311
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 166 CATGCTGGTGCAATTATTAGGAATAGCT 193
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1310 TTCCCTGGTGCAATTAGAACCAACTACCT 1283
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 4
US-10-044-090-394
; Sequence 394, Application US/10044090
; Patent No. US20020137081A1
; GENERAL INFORMATION:
; APPLICANT: Olga Bandman
; TITLE OF INVENTION: GENES DIFFERENTIALLY EXPRESSED IN VASCULAR TISSUE ACTIVATION
; FILE REFERENCE: PA-0028 US
; CURRENT APPLICATION NUMBER: US/10/044,090
; CURRENT FILING DATE: 2002-01-09
; NUMBER OF SEQ ID NOS: 850
; SOFTWARE: PERL Program
; SEQ ID NO 394
; LENGTH: 3195
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No. US20020137081A1 201572.10
US-10-044-090-394

Query Match      5.4%; Score 41.6; DB 12; Length 3195;
Best Local Similarity 52.3%; Pred. No. 0.068;
Matches 92; Conservative 0; Mismatches 84; Indels 0; Gaps 0;

QY 133 GATGGTACACAGCAAAACTTTCTGGTGTGACCATCTGGTGCATTTATTAGGAATAGC 192
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 577 GAAGATCCACTTGGACCTAATTCGTATTATGACAAAACATAATCTCTTTTGATAATAT 636
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 193 TATGTTGATTAACACTAGCAAGTGACAGCAAACTTCGTTAAGAAACAGCAGGCCAAC 252
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 637 TCTTGTGATGACATAAGAACGAGACCAACCTGGGCTGAAGAAAGAGATTAATGCT 696
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 253 TCAACCTATGACATATCTGGATCGCTCATATGCGTGGTTCCTGCAGGGCAG 308
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 697 GAACATTTGATCCCACTTCGTCCAAAACCGTGGCCGTGGGGGATACAGAGCGAG 752
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 5
US-09-850-351A-31
; Sequence 31, Application US/09850351A
; Patent No. US20020100080A1
; GENERAL INFORMATION:
; APPLICANT: Feltelson, Jerald S.
;             Schnepf, H. Ernest
;             Narva, Kenneth E.
;             Stockhoff, Brian A.
;             Schmeits, James
;             Loewer, David
;             Dullum, Charles Joseph
;             Muller-Cohn, Judy

```

```

Stamp, Lisa
Morrill, George
TITLE OF INVENTION: No. US20020100080A1el Pesticidal Toxins and Nucleotide
Sequences Which Encode These Toxins
NUMBER OF SEQUENCES: 144
CORRESPONDENCE ADDRESS:
ADDRESSER: Saliwanchik, Lloyd & Saliwanchik
STREET: 2421 N.W. 41st Street, Suite A-1
CITY: Gainesville
STATE: FL
COUNTRY: US
ZIP: 32606-6669
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/850,351A
FILING DATE: 07-May-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 09/073,898
FILING DATE: 06-May-1998
APPLICATION NUMBER: US 08/960,780
FILING DATE: 30-OCT-1997
APPLICATION NUMBER: US 60/029,848
FILING DATE: 30-OCT-1996
ATTORNEY/AGENT INFORMATION:
NAME: Sanders, Jay M.
REGISTRATION NUMBER: 39,355
REFERENCE/DOCKET NUMBER: MA-708CD1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 352-375-8100
TELEFAX: 352-372-5800
INFORMATION FOR SEQ ID NO: 31:
SEQUENCE CHARACTERISTICS:
LENGTH: 2645 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
INDIVIDUAL ISOLATE: PSL77C8
SEQUENCE DESCRIPTION: SEQ ID NO: 31:
US-09-850-351A-31

```

[illegible]

```

RESULT 6
US-09-967-805-1
; Sequence 1, Application US/09967805
; Patent No. US20020120114A1
; GENERAL INFORMATION:
; APPLICANT: Schnepf, Ernest H
; APPLICANT: Narva, Kenneth E
; APPLICANT: Stockhoff, Brian A
; APPLICANT: Finstad Lee, Stacey
; APPLICANT: Walz, Mikki
; APPLICANT: Sturgis, Blake
; TITLE OF INVENTION: Pesticidal Toxins and Genes from Bacillus laterosporus
; TITLE OF INVENTION: Strains

```

```

; FILE REFERENCE: MA-719XC2D1
; CURRENT APPLICATION NUMBER: US/09/967,805
; CURRENT FILING DATE: 2001-09-28
; PRIOR APPLICATION NUMBER: 09/371,913
; PRIOR FILING DATE: 1999-08-10
; PRIOR APPLICATION NUMBER: 60/095,955
; PRIOR FILING DATE: 1998-08-10
; PRIOR APPLICATION NUMBER: 60/138,251
; PRIOR FILING DATE: 1999-06-08
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 2645
; TYPE: DNA
; ORGANISM: Bacillus laterosporus
; US-09-967-805-1

Query Match      5.2%; Score 40; DB 10; Length 2645;
Best Local Similarity 60.4%; pred. No. 0.18;
Matches 64; Conservative 1; Mismatches 41; Indels 0; Gaps 0;

Qy      64  AAAATTACTACACATCTGACAGCTGGATTTAAATTTAAATGACGTAAGTTAGATATTAT 123
          || || || || || || || || || || || || || || || || || || || ||
Db      2460 AACGTGAGTGACACACTTGAAGTGATAAAATTTTCAAGGATGGACAATTTAAATTTGAT 2519

Qy      124  TACACAAGTGTATGGTACACAAGGACAACACTTTCTGGTGTGACCATG 169
          || || || || || || || || || || || || || || || || || || || ||
Db      2520 TTTCAAAAATATAGTAAATGAACAAGGATTTATTTATGACAGTG 2565

```

```

RESULT 7
US-09-854-133-533
; Sequence 533, Application US/09854133
; Publication No. US20020183499A1
; GENERAL INFORMATION:
; APPLICANT: Lodes, Michael J.
; APPLICANT: Mohamath, Raodoh
; APPLICANT: Henderson, Robert A.
; APPLICANT: Benson, Darin R.
; APPLICANT: Secrist, Heather
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR
; TITL OF INVENTION: THE THERAPY AND DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.475C10
; CURRENT APPLICATION NUMBER: US/09/854,133
; CURRENT FILING DATE: 2001-05-11
; NUMBER OF SEQ ID NOS: 735
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 533
; LENGTH: 289
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(289)
; OTHER INFORMATION: n = A,T,C or G
; US-09-854-133-533

```

[illegible]

; APPLICANT: Evans, Cheryl
 ; APPLICANT: Merberg, David
 ; APPLICANT: Treacy, Maurice
 ; APPLICANT: Acostino, Michael J.
 ; APPLICANT: Steininger II, Robert J.
 ; APPLICANT: Spaulding, Vikki
 ; APPLICANT: Wong, Gordon G.
 ; APPLICANT: Clark, Hilary
 ; APPLICANT: Fecthel, Kim
 ; TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES ENCODING THEM
 ; FILE REFERENCE: 6055-64X
 ; CURRENT APPLICATION NUMBER: US/09/729,674
 ; CURRENT FILING DATE: 2000-12-04
 ; PRIOR APPLICATION NUMBER: 09/539,330
 ; PRIOR FILING DATE: 2000-03-30
 ; NUMBER OF SEQ ID NOS: 283
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 141
 ; LENGTH: 2481
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 US-09-729-674-141

Query Match 4.6%; Score 35.6; DB 10; Length 2481;
 Best Local Similarity 52.7%; Pred. No. 3.2;
 Matches 77; Conservative 0; Mismatches 69; Indels 0; Gaps 0;

| | | | |
|----|-----|---|-----|
| QY | 344 | GCGGCCAGACTTCAGCCGCACGAGTTGGCGGGCACCGGTGGTGTGGCGCGTG | 403 |
| | | | |
| Dd | 402 | GCACCTGGAAGGTGACCACCAAGGTATAGAGCTGGCGGGTTCCTCGTTATAGTGGAG | 343 |
| | | | |
| QY | 404 | TTCGCGGCCAGCTGTCGGTAAAGGCTCTGTCGTGTCGATGGCCATCGGCGGTGCGG | 463 |
| | | | |
| Dd | 342 | TCTTGGCCAGAGTCCGAGAGCCGAGCATGTCCGGTGGAGTAGTCCACCACAGGCGCG | 283 |
| | | | |
| QY | 464 | TTCGTGGCGGTCTGATTGGCTCTAAA | 489 |
| | | | |
| Dd | 282 | TCAGGGCGCAGCTGATTGTCACAA | 257 |
| | | | |

RESULT 13
 US-09-875-114-2
 ; Sequence 2, Application US/09875114
 ; Patent No. US2002002131A1
 ; GENERAL INFORMATION:
 ; APPLICANT: No. US2002002131AlIthwestern University
 ; APPLICANT: No. US2002002131A1 Bouck
 ; APPLICANT: David Dawson
 ; APPLICANT: Paul Gillis
 ; TITLE OF INVENTION: Methods and Compositions for Inhibiting Angiogenesis
 ; FILE REFERENCE: 0290-23U2
 ; CURRENT APPLICATION NUMBER: US/09/875,114
 ; CURRENT FILING DATE: 2001-06-06
 ; PRIOR APPLICATION NUMBER: US 09/122,079
 ; PRIOR FILING DATE: 1998-07-23
 ; PRIOR APPLICATION NUMBER: PCT/US98/15228
 ; PRIOR FILING DATE: 1998-07-23
 ; PRIOR APPLICATION NUMBER: US 08/899,304
 ; PRIOR FILING DATE: 1997-07-23
 ; NUMBER OF SEQ ID NOS: 2
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 2
 ; LENGTH: 22484
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; NAME/KEY: Unsure
 ; LOCATION: 1....22484
 ; OTHER INFORMATION: "n" means either a, c, t, or g
 US-09-875-114-2

Query Match 4.6%; Score 35.2; DB 10; Length 22484;

Query Match 4.6%; Score 35.2; DB 10; Length 22484;

